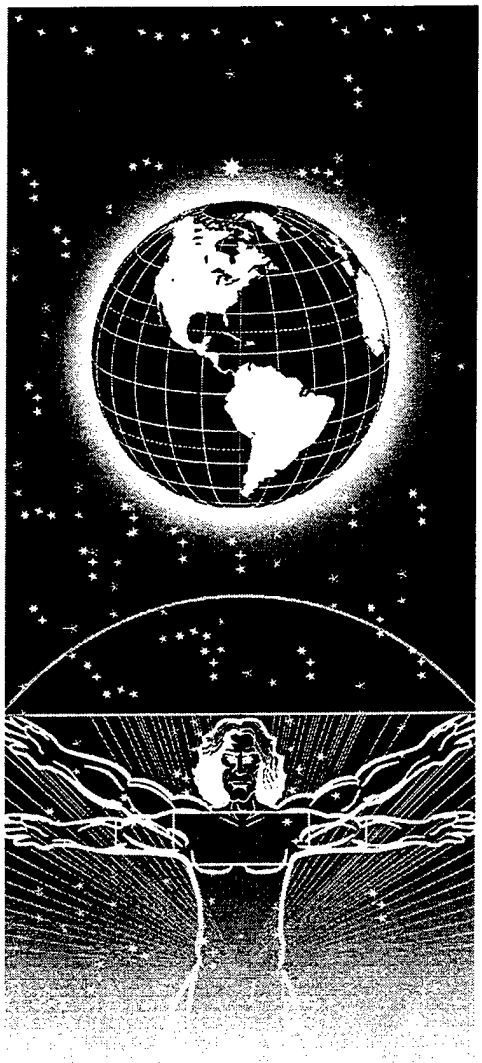
AL/OE-TR-1996-0176



UNITED STATES AIR FORCE ARMSTRONG LABORATORY

Summary of Human Health Risk Assessment Guidelines and Methodologies

Kowetha A. Davidson

Biomedical and Environmental Information Analysis Section
Health Sciences Research Division
Oak Ridge National Laboratory
Oak Ridge TN 37830

September 1996

DoD EPA DOE SERDP
Strategic Environmental Research and Development Program

A Partnership to Improve the Environment

Approved for public release;

distribution is unlimited.

Occupational and Environmental Health Directorate
Toxicology Division
2856 G Street
Wright-Patterson AFB OH 45433-7400

NOTICES

When US Government drawings, specifications or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from the Armstrong Laboratory. Additional copies may be purchased from:

NATIONAL TECHNICAL INFORMATION SERVICE
5285 PORT ROYAL ROAD
SPRINGFIELD, VIRGINIA 22161

Federal Government agencies and their contractors registered with the Defense Technical Information Center should direct requests for copies of this report to:

DEFENSE TECHNICAL INFORMATION CENTER
8725 JOHN J. KINGMAN RD STE 0944
FT BELVOIR VA 22060-6218

DISCLAIMER

This Technical Report is published as received and has not been edited by the Technical Editing Staff of the Armstrong Laboratory.

TECHNICAL REVIEW AND APPROVAL

AL/OE-TR-1996-0176

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1996	Final Report - December 1995-July 1996

4. TITLE AND SUBTITLE
Review of Human Health Risk Assessment Guidelines and Methodologies

5. FUNDING NUMBERS
Contract DE-AC05-96OR22464
SERDP CU Project #770

6. AUTHOR(S)
Kowetha A. Davidson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Biomedical and Environmental Information Analysis Section
Health Sciences Research Division
Oak Ridge National Laboratory
Oak Ridge TN 37830 (See Item 11)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Armstrong Laboratory, Occupational and Environmental Health Directorate
Toxicology Division, Human Systems Center
Air Force Materiel Command
Wright-Patterson AFB OH 45433-7400 (See Reverse)

10. SPONSORING;MONITORING AGENCY REPORT NUMBER
AL/OE-TR-1996-0176

11. SUPPLEMENTARY NOTES
Managed by: Lockheed Martin Energy Research Corporation for the U. S. Department of Energy

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This report presents an overview of risk assessment guidelines and methodologies for evaluating cancer and noncancer hazards due to exposure to environmental substances. An overview of the National Research Council's 1994 report *Science and Judgment in Risk Assessment* has been included. Although other federal agencies have established risk assessment guidelines, this report focuses on guidelines and methodologies currently in use or currently proposed by the U. S. Environmental Protection Agency. The primary purpose of this report is to identify key issues pertinent to cancer and noncancer risk assessments that impact the risk assessment process and result in more plausible risk estimates.

14. SUBJECT TERMS
Risk Assessment Guidelines, Cancer, Environmental Substances, Reviews

15. NUMBER OF PAGES
82

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Item 9, Continued:

United States Environmental Protection Agency
National Center for Environmental Assessment
26 W. Martin Luther King Drive
Cincinnati OH 45268

Navy Environmental Health Center
Deputy Director for Environmental Programs
2510 Walmer Avenue
Norfolk VA 23513-2617

Department of Energy EH-232
1000 Independence Avenue, S. W.
Washington DC 20585

Environmental Safety and Occupational Health
110 Army Pentagon, 2E577
Washington DC 20310-0110

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

PREFACE

This document was produced under the auspices of the Environmental Risk Assessment Program (ERAP), which has its genesis in the DOD/DOE Strategic Environmental Research and Developmental Program (SERDP) that was established through Public Law 101-510 (10 United States Code 2901-2904). ERAP was established as a cooperative effort of DOD, DOE, and EPA to improve health and ecological risk assessments and to foster consistency in risk assessments across federal agencies. The program has three working groups chartered under its mission which are the Materials/Chemicals Risk Assessment (MCRA) Working Group, Human Risk Assessment Methodology (HRAM) Working Group, and the Ecological Risk Assessment Methodology (ERAM) Working Group. The program also has an Advisory and Coordinating Committee (ACC) that oversees the program and the working group's activities.

This report is a product of the HRAM Working Group and presents issues concerning risk assessment guidelines and methodologies established for evaluating cancer and noncancer hazards due to exposure to environmental substances. Although other federal agencies have established risk assessment guidelines, this report focuses on the guidelines and methodologies established by the U.S. Environmental Protection Agency. Another document, *Reviews of Exposure Assessment Guidelines*, presented issues concerning exposure assessment . Therefore, these issues will not be discussed in the present document. This report was prepared by Dr Kowetha A. Davidson with help from Dr Robert A. Young (draft of the neurotoxicity sections) and Dr Carol S. Forsyth (draft of the developmental toxicity sections), all of the Oak Ridge National Laboratory, and was reviewed by members of the HRAM Working Group.

The ERAP Advisory and Coordinating Committee endorses the information contained within this document with the understanding that the end user is responsible for its application. This means that users are responsible for obtaining any internal scientific and policy reviews required prior to its acceptance within other organizations.

ENVIRONMENTAL RISK ASSESSMENT PROGRAM (ERAP)
ADVISORY AND COORDINATING COMMITTEE (ACC)

Terry Harvey, D.V.M.
U. S. Environmental Protection Agency
National Center for Environmental Assessment
26 W. Martin Luther King Drive
Cincinnati OH 45268
Telephone: (513) 569-7531
Fax: (513) 569-7475

Lt Col Terry A. Childress
OL AL HSC/OET, Bldg 79
2856 G Street
Wright-Patterson AFB OH 45433-7400
Telephone: (937) 255-3423
Fax: (937) 255-1474

CAPT David A. Macys
Deputy Director for Environmental Programs
Navy Environmental Health Center
2510 Walmer Avenue
Norfolk VA 23513-2617
Telephone: (757) 363-5565
Fax: (757) 445-7330

Mr Andy Wallo, III
Department of Energy EH-232
1000 Independence Avenue, S.W.
Washington DC 20585
Telephone: (202) 586-4996
Fax: (202) 586-3915

COL Hew Wolfe
Environmental Safety and Occupational Health
110 Army Pentagon, 2E577
Washington DC 20310-0110
Telephone: (703) 695-7824
Fax: (703) 693-8149

The members of the Human Risk Assessment Methodology Working Group are as follows:

Name	Affiliation
Dr. Kowetha A. Davidson, author	Oak Ridge National Laboratory
Dr. Marvin Frazier	U.S. Department of Energy
Dr. Richard Hertzberg*	U.S. Environmental Protection Agency
Ms. Veronique Hauschild	Department of Defense, Army
Dr. Stan Hewins	Department of Defense, Air Force
Maj. Barbara Larcom*, chairperson	Department of Defense, Air Force
Dr. Po-Yung Lu	Oak Ridge National Laboratory
Ms. Andrea Lunsford*	Department of Defense, Navy
Mr. Robert H. Ross	Oak Ridge National Laboratory
Ms. Cheryl Scott*	U.S. Environmental Protection Agency

*Technical reviewers for this report

THIS PAGE INTENTIONALLY LEFT BLANK

1. INTRODUCTION

1.1. Overview of Risk Assessment Guidelines

Regulation of chemical and physical substances in the environment is mandated by federal legislation and is implemented by the U.S. Environmental Protection Agency (EPA) and other federal agencies. The objective of the legislation is to protect human health and the environment from the possible adverse health effects of exposure to these substances. The process of evaluating the impact of various substances on health requires a risk assessment and, possibly, appropriate management of such substances.

The risk assessment is an analytical process that primarily estimates some probability of an adverse health effect occurring among human receptors when exposed to these substances. Based upon a finding of unacceptable human risk (a risk management decision), the risk management process will evaluate and apply remedial technologies to lower exposure or remove these substances and correspondingly reduce the risk to an acceptable level.

The risk assessment process has been criticized by the scientific community, industry representatives, and the lay public as not being systematic and, at times, appearing arbitrary. Additionally, decisions made in the risk assessment process may yield overly conservative risk estimates that can sometimes lead to stringent remedial objectives and unusually high remedial costs. For these reasons, guidelines for estimating the risk of exposure to environmental substances were established and they have undergone considerable changes over the years.

EPA formalized its risk assessment process in 1986 with the publication of assessment guidelines for carcinogenicity, developmental toxicity, mutagenicity, exposure to chemical mixtures, and exposure. In 1988, guidelines for female and male reproductive risks were published. The developmental toxicity guidelines were revised in 1991 and the exposure assessment guidelines in 1992. In 1993, EPA proposed guidelines for neurotoxicity risk assessment. EPA has also established methodologies for assessing risk to noncancer toxicants by route of exposure: reference dose (RfD) for the oral route (U.S. EPA, 1993) and reference concentration (RfC) for the inhalation route (U.S. EPA, 1994a). It should be mentioned that prefacing these guidelines and methodologies were the water quality criteria guidelines detailing the method for estimating acceptable daily intakes (ADI) for noncancer toxicants based on no-observed-adverse-effect levels (NOAEL) and lowest-observed-adverse-effect levels (LOAEL) and deriving slope factors (q_1^*) for carcinogens using the linearized multistage procedure (U.S. EPA, 1980). These methodologies are directly related to the present methodologies and guidelines for carcinogen risk assessment and the derivation of RfDs and RfCs for noncarcinogens.

This report will first review some of the major documents contributing to the evolution of risk assessment as practiced today; these include the National Research Council (NRC) report of 1983, the NRC report of 1994, and the Office of Science and Technology Policy report of 1985. These reports are pivotal to describing the way risk assessment is performed by regulatory agencies and are the focus on which to evaluate the risk assessment process. EPA's risk assessment guidelines will be reviewed along with published criticisms, alternatives, and recommendations to the guidelines. This is not a comprehensive report; particularly, chemical-specific issues have not been discussed. Also, the mixtures issues will not be discussed, and the exposure assessment issues are discussed in a separate report. EPA is currently working on revising the mixtures guidelines. Some subjects are discussed in detailed discussion than others, but any subject can be expanded at the request of the working group.

1.2. Risk Assessment in the Federal Government: Managing the Process (Red Book) (NRC/NAS, 1983)

Major contribution of this report: established the four steps of the risk assessment process and their contribution to the whole process.

Under a directive from congress, the U.S. Food and Drug Administration (FDA) contracted with the National Research Council's Commission on Life Sciences, which formed the Committee on the Institutional Means for Assessment of Risk, to study organizational arrangements for conducting the risk assessments in support of regulatory management by the federal government. The regulation of chemical and physical substances to which human are exposed is implemented by four government agencies: the EPA, FDA, Occupational Safety and Health Administration (OSHA), and the Consumer Product Safety Commission (CPSC). During the decade of the 70s, there was increased public concern about health hazards, (cancer and other chronic health hazards) associated with exposure toxic substances. There was also concern about the risk assessment process on which regulators based their decisions as well as the cost and benefits associated with implementing the decisions. The interaction between regulators (risk managers) and scientists (risk assessors) was cause for additional concern by scientist, industry, and the public (NRC/NAS, 1983). The objectives of the NRC's Committee on the Institutional means for Assessment of Risk (referred to as the Committee) were as follows:

1. assess the merits of separating the analytic functions of developing risk assessments from the regulatory function of making policy decisions,

2. consider the feasibility of designating a single organization to do risk assessment for all regulatory agencies,

3. consider the feasibility of developing uniform risk assessment guidelines for use by all regulatory agencies.

The Committee searched for "mechanisms to ensure that government regulations rests on the best available scientific knowledge" and to ensure that scientific integrity is maintained as science and government work together in addressing issues related to adverse effects of environmental substances on human health. The Committee noted that its conclusions and recommendations apply primarily to cancer risk associated with exposure to environmental chemicals, but they also apply to other endpoints of human health (systemic effects, developmental effects, etc.)

1.2.1. Steps in the Risk Assessment Process

The Committee established the four basic steps (hazard assessment, dose-response assessment, exposure assessment, and risk characterization) of the risk assessment paradigm, which is widely accepted today. The Committee's description of the steps and the data (information) sources is presented below:

The Committee defined *hazard identification* as the process of determining whether exposure to an agent can cause an increase in the incidence or severity of a health condition. The nature and strength of the evidence is characterized. Hazard identification answers the following question: Does the agent cause an increase in incidence or severity of an adverse effect in test animals? If the answer is yes, then the agent may pose a risk (cancer) to humans. *Data sources*: epidemiologic and other human studies (always given primary consideration when characterizing adverse health effects), animal bioassays, short-term studies, genetic toxicity, and absorption, distribution, metabolism, and absorption (ADME) data, structure-activity relationships. *Dose-response assessment* was defined as "the process of characterizing the relation between the dose of an agent administered or received and the incidence [or severity] of an adverse effect in an exposed population and estimating the incidence of the effect as a function of human exposure to the agent." What is the relationship between the dose and incidence or severity of the adverse effect? *Data sources*: human data are given primary consideration, if available, when characterizing adverse health effects, but quantitative data from human studies are usually not adequate; animal data are usually available, but species extrapolation is required. In addition, high occupational exposures of humans are extrapolated to low exposures to the general population or high experimental doses in animals are extrapolated to low doses is required.

The definition given for *exposure assessment* was "the process of measuring or estimating the intensity, frequency, and duration of human exposures to an agent currently present in the

environment or estimating hypothetical exposures that might arise from the release of new chemicals into the environment." What exposures are currently experienced by the population or anticipated to occur under different conditions? Exposure assessments describes the various aspects of exposure including intensity, duration, route, frequency, populations, and uncertainties associated with the estimates. *Data sources*: direct measurements of chemicals in environmental media, models that predict exposures in environmental media, extrapolations from small segment to large general populations.

Risk characterization is the process of estimating the incidence of a health effect under the various conditions of human exposure described in exposure assessment. What is the estimated incidence or severity of the adverse effect that would occur in a given population? Risk characterization was described as a summary of the dose-response and exposure assessments and associated uncertainties. *Data sources*: hazard assessment, dose-response assessment, and exposure assessment provide the information for making predictions for different population groups.

1.2.2. Components in the Risk Assessment

The NRC Committee compiled a list of components in the risk assessment process that arise from attempts to bridge the gap between inherent uncertainties (missing or ambiguous data or gaps in current scientific theory) and the need to conduct the assessment. The list compiled by the Committee was not exhaustive, but 25 components were associated with hazard assessment and 13 with dose-response assessment. These components focused primarily on cancer assessments but can be extended to evaluate other adverse effects in humans. The following section includes a condensed versions of these components.

1.2.2.1. Components of hazard identification – The components related to performing a hazard assessment are listed below (organized according to the data sources of a hazard assessment).

Epidemiologic studies:

- the relative weight given to different types of studies or to studies with different results
- the level of statistical significance required for positive results
- the significance of positive findings in studies in which route of exposure is different from the one of interest
- how to combine different types of responses

Animal studies:

- the level of confirmation required for positive results (two or more studies?); weight given to negative studies
- the quality and statistical power of the studies as the basis for weighing studies
- the handling of differences in metabolism and pharmacokinetics between animals and humans and the incorporation of the differences in the results
- the weight given to rare tumors, especially when the incidences are not statistically significant
- the weight given to studies when tissue damage or other toxic effects accompany a carcinogenic effect
- the combining of benign and malignant tumors
- the weight given to decrease latency

Short-term studies:

- weight given to short-term results
- level of evidence required for addition to weight of evidence
- weight given to different types of tests
- weight given positive versus negative results

Structural-activity analysis:

- weight given to results with structurally similar compounds.

1.2.2.2. Components of Dose-Response Assessment – The components related to the dose-response assessment are listed below. The data sources for dose-response assessment are epidemiologic (human) studies and animal bioassays. The components associated with human data are encountered only when these data are available, which in not often. The two major components of dose-response assessment are extrapolation from high to low doses and interspecies dose conversion. These components focus primarily on cancer assessments, but can be applied to other adverse effects in humans.

Epidemiologic studies:

- selection of a dose-response model to extrapolate to low environmental doses
- Selection of best estimate of risk or upper confidence limit of risk
- method for adjusting for comparatively short follow-up period in epidemiologic studies
- health effects on which to derive estimates; for example should consideration be given only to cancers unequivocally related to exposure or all types of cancer
- how to account for exposures to other potential carcinogens
- how to account for differences in the temporal pattern of exposed population and the population in question (lifetime risk function of total dose no matter when it was received during the lifetime; weight given to the most recent exposure)
- how to account for possible physiological differences between exposed population and population in question.

Animal data:

- selection of mathematical models used to extrapolate from experimental animal doses to environmental human exposures
- Selection of best estimate of the risk or the upper confidence limit of the risk
- the dose scaling method for converting animal doses to human doses
- incorporation of metabolism data into the assessment
- how data on more than one nonhuman species or strain be used in the assessment; use most sensitive species or strain; combine data on different species and strain and method for combining
- how to use data on more than one tumor type; combine data or use tumor type most affected by exposure
- interpretation of statistically significant decreases in tumor incidences at specific anatomical sites.

The Committee noted that choices made for each component when missing or ambiguous data are encountered are called inference or "default options". Inference or default options are based on both scientific and policy judgements. Science and policy judgments in risk assessment can determine the outcome of the assessment, because the judgements determine the defaults options chosen in a risk assessment, and the choices of default options affect the conservativeness of the assessment. If policy judgments are the basis for choosing defaults, then the policy should be grounded, as much as current advancements allows, in scientific knowledge. The committee

further noted that, when conservatism drives the choices for default options, the risk assessment policies may be driven more by risk management considerations than by science. Risk management takes into account the non-science aspects (political, social, economic, and other considerations) of regulating chemical exposures. In addition, within the framework, there should be clear separation of default options based on scientific judgement and those based on policy.

The Committee also described the need to establish guidelines that direct the risk assessor in conducting the assessment. Guidelines allow the separation of risk assessment and risk management by laying down a formal procedure for conducting risk assessments. Guidelines would aid in the quality control over risk assessments, ensuring that the assessment conforms to the scientific judgements of experts in the diverse fields encompassing risk assessments. Guidelines also ensure that the assessments are clear, complete, and comprehensive. Additionally, the guidelines ensure that risk assessments are consistent and predictable for different chemicals. The Committee noted that the guidelines must be comprehensive in detail, but not so inflexible as to limit scientific interpretation of data or change of a default option in the face of reliable scientific evidence. The Committee stated that uniform guidelines for all aspects of risk assessment, except exposure assessment, are feasible and desirable for governmental agencies.

1.3. Chemical Carcinogens: Review of the Science and its Associated Principles (OSTP, 1985)

Major contributions: reiterated the risk assessment paradigm established by NRC/NAS (1983) and listed principles on which to build risk assessments.

The Office of Science and Technology Policy (OSTP) presented its general scientific view of carcinogenesis and general principles on which federal agencies tailor their risk assessment guidelines to meet legislative requirements. OSTP divided risk into two parts: hazard and exposure. Hazard being the toxicity deduced from a variety of studies of humans or animals, and exposure being the contact of individuals with a substance. There were a total of 31 principles put forth by OSTP; all are not listed below, particularly those related to exposure assessment.

1. Carcinogenesis is a multistage process involving direct or indirect effects on the genome; the process may be influenced by a number of factors, such as age, sex, diet, hormonal status, and genetic background.

2. Carcinogenicity can be influenced by induction of nonphysiological responses (excessive organ damage, hormonal disruption, metabolic saturation, etc) that may affect the relevancy of the test system for evaluating human carcinogenicity.

3. Mechanistic considerations, such as DNA repair or damage, do not prove or disprove the existence of a threshold for carcinogenesis.

4. Short-term tests, particularly genotoxicity tests, are useful in providing information for interpreting carcinogenicity studies, but are limited in their ability to predict carcinogenicity, and thus, cannot substitute for long-term animal or epidemiologic studies.

5. The statement,"In the absence of adequate data in humans, it is reasonable, for practical purposes, to regard chemicals for which there is sufficient evidence of carcinogenicity in animals as it they presented a carcinogenic risk to humans", is a reiteration of the IARC principal. Other relevant information must be considered in reaching a conclusion based on long-term animal studies.

6. Animal models having high background tumor incidences pose special problems, and such data must be interpreted carefully.

7. Long-term animals studies should achieve adequate biological and statistical sensitivity and adequate biological and statistical specificity to avoid producing false negatives and false positives. High doses for animal studies are required to achieve statistical significance, but doses should be compatible with normal life span (except that due to cancer) and minimal organ toxicity.

8. Evidence of carcinogenicity should consider relevant biological and biochemical data.

9. Evidence of probable reproducibility of long-term studies (independent confirmation of results), evidence of a dose-response, increased tumor incidence at multiple sites, and decreased tumor latency increase confidence in the study.

10. Biological plausibility of a neoplastic response may be increased when the incidence of the corresponding preneoplastic lesions is increased.

11. Well-designed and well-conducted cohort or case-control epidemiologic studies can provide data for causal association of exposure with cancer in humans.

12. Well-designed and well-conducted negative epidemiologic studies, while useful, cannot prove the lack of an association between exposure and human cancer.

13. The exposure routes in animal studies should be comparable to human exposure routes.

14. Evaluations on carcinogenicity should be based on relevant data, whether it comes from animal studies, epidemiologic studies, *in vitro* or *in vivo* short-term tests, metabolic and pharmacokinetic studies, mechanistic studies, or structure/activity analysis.

15. Mathematical models for low-dose extrapolation should be consistent with the evidence, but when evidence is limited, low-dose linearity is the preferred method.

16. Quantification of uncertainty is an important issue in risk estimation; sources of uncertainty include the model selected for low-dose extrapolation, statistical uncertainty associated with the risk estimate, and the use of animal models as test organisms.

17. Clear distinctions should be maintained among facts (statements supported by data), consensus (statements generally held by the scientific community), assumptions (statements made to fill data gaps), and science policy (statements made to resolve points of current controversy).

18. Because the human population varies in their susceptibility to chemical exposures, consideration should be given to identifying high risk groups.

1.4. Science and Judgment in Risk Assessment (NRC/NAS, 1994)

Major contributions of this report: reevaluated EPA's risk assessment guidelines in light of the Clean Air Act Amendment of 1990; proposed that EPA define default options and establish principles for moving beyond default options,

modify its hazard assessment classification, and establish a more rigorous uncertainty analysis in risk assessments.

The U.S. Congress charged the National Research Council (NRC) with the following tasks:

1. review the methods used by EPA in determining the carcinogenic risk associated with exposure to hazardous air pollutants,

2. include in its review, evaluation of the methods used for estimating the carcinogenic potency of hazardous air pollutants and for estimating human exposures to these air pollutants, and

3. evaluate, to the extent practicable, risk-assessment methods for noncancer health effects for which safe thresholds might not exist.

To perform this task, the NRC established the Committee on Risk Assessment of Hazardous Air Pollutants consisting of 25 members representing various disciplines. The Committee evaluated EPA's current risk assessment practices and noted that EPA generally followed the recommendations of the 1983 NRC report, but as new information has become available over the years, criticisms of EPA's practices have come from various groups including industry, environmental organizations, and academia. The major criticisms of EPA's risk assessment practices have been related to the lack of quantitative data, the different scientific interpretations of pertinent data, the level of uncertainty, and the incorporation of conservative default options. The committee addressed in detail six issues of EPA's risk assessment process: default options, data needs, validation, uncertainty, variability, and aggregation.

"Default Options": The Committee was concerned that EPA did not clearly identify all its "default options", nor did EPA fully explain the basis for default options. Further, EPA allows departure from default options, but has not identified the criteria for the departures. The Committee recommended that EPA continue to use default options as a means for dealing with "uncertainty about underlying mechanisms in selecting methods and models for use in risk assessment." However, EPA should identify each use of a default option and present the scientific and policy basis for the default option. The Committee also stated that EPA should formalize in its guidelines, principles for departing from default options, so as to prevent *ad hoc* undocumented departures that could damage the credibility of the assessment. The Committee identified the following objectives that EPA should consider in establishing its default options and principles for departure: protecting the public health, ensuring scientific validity, minimizing serious errors in estimating risks, maximizing incentives for research, creating an orderly and predictable process, and fostering openness and trustworthiness. The Committee, however, could not agree on which principles EPA should adopt.

Methods, Models, and Validation: The Committee expressed concern that the predictive accuracy and uncertainty in the models (exposure and toxicity assessments) EPA uses in its risk assessments are not always clearly explained or understood. The Committee noted that EPA classifies potential carcinogens based on the strength-of-evidence [should have said weight-of-evidence] associated with levels of evidence ("sufficient", "limited", "inadequate", or "no data") achieved from human and/or animal studies; the levels of evidence are grouped into A, B, and C categories based on a combination of evidence levels for humans and animals. The Committee criticized EPA's classification scheme because chemicals showing strong evidence for carcinogenicity in humans (class A) could pose a low risk due to low exposure or potency, whereas a chemical showing strong evidence only in animals (class B2) may pose a high risk due to high exposure or potency, yet the class A chemical may be viewed as a greater hazard (The Committee referred to this situation as "accidents of fate"). The Committee failed to note that it is likely that the basis for the A classification is due to cancer mortality (a less sensitive endpoint than diagnosis of cancer for estimating risk); thus, the risk of dying from cancer is an inherent underestimation of the risk of developing cancer. The Committee recommended a scheme using four categories, each of which had two to four subcategories and a descriptive narrative. The descriptive narrative would include "relevance information" based on the animal model and exposure.

The Committee also noted that EPA uses the linearized multistage model (default model) for extrapolating from human occupational exposures or experimental animal doses to low exposures for human populations to estimate the carcinogen potency of a substance. The potency is based on upper bound estimates of the risk. The Committee recommended that EPA include data on mode of action in its quantitative models and that pharmacokinetics data be incorporated into its models to extrapolate from animal to human doses, extrapolate between routes, and to link exposure to dose. The Committee recommended that EPA validate models used in risk assessments. The Committee further recommended that EPA continue to use the linearized multistage model as its default options for extrapolating to low doses, but EPA should develop criteria for using alternative models. The upper bound estimate should continue to be used to estimate the risk for developing cancer due to lifetime exposure.

The report noted that EPA uses the NOAEL/LOAEL approach for establishing safe doses for noncarcinogen effects. The Committee recommended that EPA should continue to explore the use of pharmacokinetic models for establishing target tissue doses and biologically-based quantitative models for linking exposure and noncancer effects.

Data needs: The Committee recommended that EPA develop a two-level plan for risk assessments: screening and full risk assessment. A screening assessment would require only a minimal data set, whereas a full risk assessment would require a rich data set. The Committee suggested requirements for gathering and assessing toxicologic data for the screening and full risk assessments. Generic and acute toxicity data should be collected on all chemicals as a starting point; and toxicokinetic, genetic, subchronic and chronic animal data, human toxicity data, and mechanistic data should be collected on chemicals for which there is a cause for concern. Other factors (emissions, environmental fate and transport, and exposure data) in addition to toxicity determine the level of priority for conducting a full risk assessment. The Committee recommended that data gathering and assessment for either a screening or full scale risk assessment be an iterative process.

Variability: According to the Committee, EPA has not addressed the issue of variability (age, sex, race, ethnicity, lifestyle, etc.) in its cancer risk assessment guidelines. The Committee, therefore, recommended that EPA sponsor research to study variability in susceptibility to cancer, adopt default options to account for differences in susceptibility, validate or improve the default assumption that all humans have the same susceptibility as those in epidemiologic studies and/or the most sensitive animals tested, assess risk to infants and children when their risk appear to be greater, clearly state default assumptions for nonthreshold low-dose linearity of genetic effects, and maintain a distinction between uncertainty and variability.

Uncertainty: The Committee noted that EPA's current approach to uncertainty analysis is a qualitative description of the model uncertainty rather than a quantitative analysis of model parameters. The Committee recommended that EPA develop a formal process of uncertainty analysis as part of its risk assessment process, include an analysis of other models that may be more "accurate", and present risk managers a range of risk values rather than a single point estimate.

Aggregation: Risk assessment usually addresses the hazards and risks associated with single chemical exposures. However, populations or individuals are exposed to multiple chemicals by more than one pathway. In addition, bioassay data often reveal that tumors develop at more than one anatomical site. The Committee recommended that EPA use appropriate statistical procedures to aggregate exposures to multiple chemicals and add individual potency estimates for each relevant tumor types in cases of multiple anatomical targets.

2. OVERVIEW OF EPA's 1986 CANCER RISK ASSESSMENT GUIDELINES

EPA's 1986 guidelines appear to be an extension of the 1980 guidelines for deriving water quality criteria based on threshold and nonthreshold effects (U.S. EPA, 1980). The water quality criteria guidelines first introduced the linearized multistage model as a procedure for low-dose extrapolation of cancer incidence data. The water quality criteria guidelines also provided default practices for interspecies scaling, estimating internal doses from feeding and inhalation studies, and adjusting potency estimates when the duration of the study is less than the theoretical life span of the species. EPA's 1986 carcinogen risk assessment guidelines introduced the four step paradigm identified by the NRC/NAS (1983) as the foundation of its risk assessment process (U.S. EPA, 1986). Two steps of this process, hazard identification and dose-response assessment, will be discussed in this report.

2.1. Hazard Identification

Hazard identification (qualitative aspect of carcinogen risk assessment) consists of an evaluation of the pertinent data to establish the link between exposure to a substance and adverse effects or hazards in humans. The data sources listed by EPA include epidemiologic and other human studies, animal toxicologic studies (particularly long-term exposure studies), genetic toxicity studies, short-term (or subchronic) toxicity studies relevant to cancer, metabolism and pharmacokinetics studies, and physicochemical data (U.S. EPA, 1986). The evidence from human and long-term animals studies is evaluated based on strengths and weaknesses of the studies with the evidence classified as "sufficient", "limited", or "inadequate", depending on the level of causality determined from human and animal studies separately. Evidence from other types of studies may add to or subtract from the weight of evidence.

Hazard identification concludes with an overall weight of evidence consisting of a short narrative on the strength of evidence and a classification based on the A, B, C scheme, similar to that of the International Agency for Research on Cancer (IARC). The five weight-of-evidence categories as established by EPA are: Group A: sufficient evidence from human studies, any level of evidence from animal studies; Group B: limited evidence from human studies and any level of evidence from animal studies (B1) or sufficient evidence from animal studies and inadequate of no evidence from human studies (B2); Group C: no data or inadequate evidence from human studies and limited evidence from animal studies; Group D: no data or inadequate evidence from human and animal studies; and Group E; evidence of no carcinogenicity from human or animal studies.

2.2. Dose-Response Assessment

According to the 1986 guidelines (U.S. EPA, 1986), *dose-response assessment* (quantitative aspect of risk assessment) uses cancer incidence data from humans or animals studies and exposure data (a presumed surrogate for dose) to estimate an upper bound on risk (q_1^* or slope factor) using the linearized multistage model (default model). The multistage procedure can be used to estimate risk in the experimental dose range or at low dose levels. Models other than the linearized multistage may be used for low-dose extrapolation if the data suggest that it may be more plausible. Most epidemiologic data are obtained from studies of humans exposed to high concentrations of a substance in occupational environments; thus, low-dose extrapolation would be necessary to estimate the risk to humans at low environmental exposures. For most substances, however, quantitative human data are not available; thus, interspecies and low-dose extrapolations are necessary to estimate potential cancer risk to humans. For interspecies extrapolation, animals doses are scaled to human equivalent doses, averaged over an entire lifetime (70 years (default value for humans), and expressed as a daily dose. According to the 1986 cancer guidelines, animal doses are scaled to human equivalent doses based on surface area expressed as of the ratio of the body weights to the ⅔-power (default); this practice was described in EPA's water quality criteria guidelines (U.S. EPA, 1980).

When several data sets are available, EPA's carcinogen risk assessment guidelines (U.S. EPA, 1986) state that all relevant animal data should be evaluated for quantitative risk estimations, but the most emphasis should be placed on studies showing the greatest sensitivity (default position). However, due regard should be given to statistical and biological considerations in choosing this approach. When the environmental route of exposure is different from that of the dose-response data, route-to-route extrapolation is conducted in accordance with existing pharmacokinetic and metabolism data on the chemical. When more than one tumor site shows statistically significant elevated incidences, the individual incidences can be pooled for risk estimation (default position). The guidelines also relied on a qualitative description of the uncertainties associated with quantitative estimates of risk assessment.

3. OVERVIEW OF EPA's 1996 CANCER RISK ASSESSMENT GUIDELINES

The current external review draft of EPA's proposed carcinogen risk assessment guidelines is the result of several workshops sponsored by EPA, a workshop sponsored by the Society of Risk Analysis, a review by the Office of Science and Technology Policy, and the NRC's comments in *Science and Judgment in Risk Assessment* (U.S. EPA, 1996b) The gradual evolution of the 1986 guidelines into the 1996 proposed guidelines has produced several key changes in the hazard identification and dose-response assessment phases of the carcinogen risk assessment process. These changes will be discussed below.

3.1. Hazard Assessment and Characterization

The purpose of hazard assessment is to present and evaluate pertinent data to determine whether an agent poses a carcinogenic hazard to humans and under what circumstances the hazard may be expressed. The assessment consists of an evaluation of all pertinent data, not just human and animal cancer data, to arrive at a conclusion regarding the carcinogenicity of an agent. In addition to the traditional assessment, which included human and animal cancer data, toxicokinetic, metabolism, and an analysis of structurally related compounds, an analysis of mode of action has become an important aspect of the proposed hazard assessment and characterization process.

Mode of action information provides insight into the relevance of animal data to human carcinogenicity, the conditions under which carcinogenicity may be expressed in humans, and the selection of a dose-response approach. Questions to be answered by analyzing the mode of action include:

- Does the agent affect DNA directly or indirectly?
- Does the agent affect cell proliferation, apoptosis, gene expression, immune surveillance, or other cellular mechanisms not involving DNA?
- Does the agent act by a mode of action reasonably anticipated to occur in humans or by one known not to occur in humans.

Animal data often provide clues as to possible modes of action. For example, agents that induce tumors at multiple site and in multiple species are likely to be mutagenic, whereas agents that affect tumors with high spontaneous incidence rates or induce only late-developing benign tumors suggests a growth promoting mode of action. Therefore, all data related to the mode of action should be evaluated very carefully and incorporated into the weight of evidence analysis.

The weight of evidence analysis uses the entire body of evidence to make a sound judgement as to the potential carcinogenicity of the agent for humans. Consideration is given to quality, consistency, and volume of data. The most weight is given to multiple well-conducted human and animal studies showing consistent responses across studies combined with strong data sets on toxicokinetic, metabolic fate, mode of action, structural activity relationship, and other key evidence such as physicochemical properties. The entire evaluation of all key data elements are combined to develop a conclusion regarding the carcinogenicity of an agent. EPA listed several factors for weighing the totality of evidence.

- Evidence of human causality
- Evidence of animal effects relevant to humans
- Coherent inferences
- Comparable metabolism and toxicokinetics between species
- Mode of action comparable across species

Decreases in the strength of evidence in these categories can result in decrease weight as to the carcinogenicity of an agent to humans.

The proposed guidelines do not categorize substances within the A,B,C weight-of-evidence groups. EPA has established three categories of descriptors for human carcinogenic potential: "*known/likely*", "*cannot be determined*", and "*not likely*." The categories are proposed to be route specific. The *known* category is used when a definite causal association between human exposure and cancer can be established based human data. A subcategory "treated as if they were *known*" human carcinogens is used for agents for which human evidence is not strong enough to show a definite causal association, but experimental animal evidence is strong. The *likely* category is used when strong animals data show carcinogenicity by a mode of action considered to be relevant or assumed to be relevant to human carcinogenicity; human evidence in this case may be weak or absent. Two subcategories to *likely* are *high end* of the totality of evidence and *low end* of the totality of evidence. In the latter subcategory, the evidence is decidedly weak, but still showing potential evidence of carcinogenicity to humans. The *cannot be determined* category is used when the evidence of potential carcinogenicity is based on data that are *suggestive*, *conflicting*, or *inadequate*; it is also used when *no data* are available to perform an assessment. The *not likely* category is used (1) when more than one well-conducted study in at least two appropriate species show no evidence of carcinogenicity, (2) when the mode of action is not relevant to humans, (3) for a particular route when evidence shows route specificity, (4) for a specific dose when evidence

shows dose limitations, and (5) when extensive human experience shows no evidence of a carcinogenic effect.

The weight of evidence analysis is followed by the hazard narrative, which summarizes the results of the hazard assessment. The narrative explains the likelihood of hazard in humans, the conditions under which the hazard would or would not be expressed, strengths and weaknesses of the evidence, the mode of action, and the impact of the mode of action on dose-response assessment.

3.2. Dose-Response Assessment

The dose-response assessment proceeds in two steps. The first step is to model the data in the range of experimental observation; the second step is to extrapolate below the experimental range to lower environmental exposures using specific models or default procedures. When data are available, biologically-based or case-specific models are used for both the experimental range and extrapolation to low doses. However, rarely will there be sufficient data or resources to employ either biologically-based or case-specific models for dose-response assessments. Instead, curve-fitting models are used for dose-response assessment of animal data in the range of experimental observation and estimating the point of departure for extrapolating to low doses. The selection of the curve-fitting model should be consistent with data for which it is applied. The point of departure is the lower 95% confidence limit on dose associated with 10% extra risk (LED_{10}). The point of departure (LED_{10}) is a matter of science policy that will result in consistency among different assessments and consistency with the benchmark approach for noncancer assessments. When sufficient data are available, the point of departure may be set below the LED_{10}. Comments are being sought on using others points on the dose-response curve for the point of departure. Modeling of human data are conducted on a case by case basis.

If data are not available for applying biologically based or case-specific models, one of three science policy default procedures is selected for extrapolation of doses below the point of departure. The three procedures are linear, nonlinear, and both. The mode of action is a primary factor in selecting the default procedure.

The default linear procedure involves straight line extrapolation from the point of departure to the origin or zero response. When the evidence suggests a mode of action involving gene mutation due to DNA reactivity or another mode action anticipated to be linear, a linear procedure is selected. A linear procedure is also selected when there is insufficient evidence for applying either the linear or nonlinear procedure.

When there is sufficient evidence to support a nonlinear dose-response or evidence of a threshold response and no evidence for a linear dose-response, a nonlinear procedure is selected for the dose-response assessment. Probabilistic dose-response functions are not fitted to nonlinear dose-response data for low-dose extrapolation; instead the margin of exposure is calculated (science policy). The margin of exposure is the point of departure, usually the LED_{10}, divided by the environmental exposure of interest. If the dose-response data suggest a threshold response, as seen when carcinogenicity is a secondary response to toxicity or cell proliferation, the margin of exposure procedure is conducted similar to what procedure is used for noncancer endpoints including estimating an RfD or RfC for the primary effect. The guidelines proposed that factors accounting for intraspecies differences (accounting for human variability) and interspecies differences (accounting for differences in sensitivity between humans and test species) can be employed in the margin of exposure analysis. In deciding what constitutes an acceptable margin of exposure the following factors should be considered: (1) slope of the dose-response curve at the point of departure; (2) nature of the response, tumors, frank toxicity, or precursor effect; (3) nature and extent of human variability in sensitivity; (4) persistent of the agent in the body; and (5) human sensitivity compared with that of animals.

When data are available to support both linear and nonlinear procedures, as may be the case for multiple tumor sites, appropriate dose-response procedures are applied.

The selection of data for dose-response assessment first considers the positive quantitative data from well-designed, well-conducted epidemiologic studies. If adequate human data are not available, then priority is given to data from the animal species showing the greatest similarity to humans. If this cannot be determined, the quantitative evaluation should consider all animal data and base the risk estimate on the data set best representing the response in humans. Biological plausibility and the mode of action should receive major consideration when deciding on the data set(s) to be incorporated into the risk assessment. The risk estimate may be the result of (1) a single data set, (2) combined data sets from different experiments, (3) a range of estimates from several data sets, (4) pooled data sets from a single experiment, (5) an analysis of different data sets based on different modes of action, or (5) a combination of the above.

The default measure of exposure to carcinogens is the cumulative lifetime dose expressed as the average daily dose. The dose data applied to the dose-response assessment may be estimated from human studies or from animal experiments. Nevertheless, decisions have to be made on whether to use applied dose, internal dose, or delivered dose to the target organ and whether the doses are to be expressed in terms of parent compound or metabolite. Estimation of

doses from human studies are conducted on a case-by-case basis. If dose estimates from human studies are not available, the preferred method is to use toxicokinetic and toxicodynamic data to estimate human equivalent doses from animal experimental doses. Toxicokinetic and toxicodynamic data can be used to build agent-specific models for scaling internal or delivered animal doses to equivalent human doses. These models require comprehensive data sets, which are seldom available.

In the absence of toxicokinetic data, doses for oral exposure are estimated using the default procedure for interspecies scaling, in which the daily dose applied over a lifetime is scaled proportionally to the ¾-power of the body weight ($BW^{¾}$). Documentation for scaling daily doses based on $BW^{¾}$ was published in a Federal Register report (U.S. EPA, 1992b). For inhalation exposure, default methods for estimating human equivalent concentrations for particles and gases are described in the methodology for deriving RfCs (U.S. EPA, 1995).

Quantitative route-to-route extrapolation may be conducted when the route of interest is not the same as the route for which data are available. In the absence of data to the contrary, EPA's default assumption is an agent that causes internal tumors by one route may be carcinogenic by another route if the substance is absorbed by a second route to give an internal dose. Therefore, route extrapolation should be supported qualitatively before attempting quantitative extrapolation. The site of tumor formation (should be distant from the portal of entry), absorption similarities or differences between the two routes, and available toxicokinetic and toxicodynamic data must be evaluated for each case. Regardless of the qualitative support for route-to-route extrapolation, quantitative extrapolations are problematic because of first pass effects, which may alter biological responses.

Toxicity equivalence factors (TEF) are sometimes used to derive quantitative risk estimates for agents within classes of compounds. One class member serves as the reference by which other members are indexed according to shared characteristics. Although the guidelines did not refer to TEFs as screening risk values, they should be considered as such, because TEF are replaced with defensible values when sufficient data become available. So far, dioxins and furans are the only classes of compounds with adequate data to support TEFs. Criteria for developing TEFs have been presented by EPA (U.S. EPA 1991a) and are being developed and expanded in the revision of the mixtures guidelines (ILSI, 1996, in preparation).

The dose-response assessments concludes with a dose-response characterization that describes the judgments and rationales made in selecting the approach for the analysis. Plausible

alternative approaches may be presented, but the preferred approaches should be described. The uncertainties are described and quantified where practicable. Two types of uncertainties are usually encountered, model uncertainty and parameter uncertainty. Model uncertainties are not amenable to quantitation and are described qualitatively. Parameter uncertainties are described quantitatively using sensitivity analysis and statistical uncertainty analysis. Dose-response estimates are presented to one significant figure, along with an indication of whether the values are upper bounds or central tendency. Additionally, the characterization should include a discussion on likely overestimation or underestimation of the results.

3.3. Default Assumptions

In response to the NAS/NRC (1994) report, the EPA addressed the basis and justification for using default assumption in hazard and dose-response assessments. The guidelines addressed the issues and major default assumptions common to risk assessments. The major default assumptions are described within the framework of the following questions. Some of these issues will be discussed in further detail in Section 5 (Issues)

- **Is the presence or absence of effects observed in a human population predictive of effects in another exposed human population?**

Human data are typically obtained from occupational exposure studies, where the working population is different from that of the general population by sex, age, and general health and is not representative of the general population. EPA has two default assumptions concerning this issue. The first is that *when cancer effects in exposed humans are attributed to exposure to an exogenous agent, such data are predictive of cancer in any other human population exposed to the same agent.* This default is not considered to be public health conservative. EPA (U.S. EPA, 1996b) further states that when specific data on sensitive subpopulations are available, these data should be used in the assessment. The different types of susceptibilities are quite numerous and may be related to age (children may be more susceptible that adults), sex (male/female differences), ethnic or racial background, nutritional status or diet, and genetics (e.g., slow and rapid acetylators). Pertinent questions related to this issue are: How do we identify and quantitate the number or proportion of susceptible individuals in a population? How do we quantitate the cancer risk of a susceptible population? How is this information incorporated into the hazard and dose-response assessments and characterizations?

Because null results maybe obtained from studies of worker populations, the second default assumption is as follows: *When cancer effects are not found in an exposed human population, this information by itself is not generally sufficient to conclude that the agent poses no carcinogenic*

hazard to this or other human populations exposed to the same agent. How much weight can one place on well-conducted epidemiologic studies producing null results? Should these results carry more or less weight than positive animal studies? According to the 1996 proposed cancer guidelines, studies of worker populations producing null results may not have the power to detect effects in sensitive population, suggesting that null results carry very little weight in hazard assessments. The 1996 proposed guidelines also stated that quantitative data obtained from null results may be used to estimate upper bounds on human risk for comparison with estimates from animal data.

- **Is the presence or absence of effects observed in an animal population predictive of effects in exposed humans?**

The default assumption states that *positive effects in animal cancer studies indicate that the agent under study can have carcinogenic potential in humans.* This assumption is considered to be public health conservative. This default assumption has several subparts that are addressed by additional default assumptions. (1) *Effects seen at the highest dose tested are appropriate for assessment, but it is necessary that the experimental conditions be scrutinized.* To improve the detection power of animal studies, high doses are used that often cause toxicity manifested by effects such as cell killing and compensatory cell proliferation. If excessive toxicity is observed, a study may be discarded, but expert judgement is required in this decision process. (2) *Target organ concordance is not a prerequisite for evaluating the implications of animal study results for humans.* This is a public health conservative science policy option. The carcinogenic targets for animal and human studies may or may not be concordant. Nevertheless, data that do not support site concordance should be considered when available. Site concordance is inherently assumed when toxicokinetic modeling is used to estimate target doses for humans. (3) *Include benign tumors observed in animal studies in the assessment of animal tumor incidence if they have the capacity to progress to the malignancies with which they are associated.* This is a science policy decision more public health conservative than the alternative. (4) *Benign tumors that are not observed to progress to malignancy are assessed on a case-by-case basis.* It should be noted that animal studies are usually terminated after 2 years, but humans receive continued exposure until death. Also some tumors do not progress because of a short duration of exposure, but they would probably progress if exposure had been continued for a longer time.

- **How do metabolic pathways relate across species?**

The default assumption is that *there is similarity of the basic pathways of metabolism and the occurrence of metabolites in tissues in regard to the species-to-species extrapolation of cancer hazard and risk.* There may be quantitative differences in metabolic pathways; unless data show

qualitative differences, metabolism is assumed to be similar in animals and humans.

- **How do toxicokinetic processes relate across species?**

The default option for oral exposure is that *a human equivalent dose is estimated from data on another species by an adjustment of animal oral dose by a scaling factor of body weight to the 0.75 power* ($BW^{0.75}$). Scaling based on body weight makes the assumption that the area under the curve (AUC) is equivalent across species for dosimetric purposes. This value is based on scaling of metabolic processes across species of different sizes. The default for inhalation exposure is that *a human equivalent dose is estimated by default methodologies that provide estimates of lung deposition and of internal dose.* As new data becomes available, both defaults can be replaced. The default for route-to-route extrapolation is that *an agent that causes internal tumors by one route of exposure will be carcinogenic by another route if it is absorbed by the second route to give an internal dose.* This is a public health conservative default option that assumes qualitative similarity of metabolic processes across routes of exposure; adequate data are required for route specific designation.

- **What is the correlation of the observed dose-response relationship to the relationship at lower doses?**

Biologically-based or case-specific models are used when sufficient data are available. In the absence of sufficient data the default procedure is to *use a curve-fitting model for the observed range of data when the preferred approach cannot be used.* There are three default procedures to consider when extrapolating dose-response data to low doses (linear, nonlinear, and both). Mode of action is a primary feature in selecting the best approach to use. *A linear default approach is used when the mode of action information is supportive of linearity or there is insufficient data to support a nonlinear mode of action.* For linear extrapolation, a straight line is drawn from the point of departure to the zero response or zero dose. The default point of departure is the LED_{10} (public health conservative) when data are not available to support a lower point of departure. The straight line extrapolation gives an upper bound on risk at low doses. *When adequate data on mode of action show that linearity is not the most reasonable working judgement and provide sufficient evidence to support a nonlinear mode of action, the default changes to a margin of exposure analysis, which assumes that nonlinearity is more reasonable.* The point of departure is the LED_{10} unless data are available to support a lower point. *Both linear and margin of exposure procedures can be used when the mode of action data indicate that the dose response may be adequately described by both linear and nonlinear approaches.*

4. OVERVIEW OF NONCANCER RISK ASSESSMENT GUIDELINES

In 1980, EPA published guidelines for deriving water quality criteria for threshold toxicants (U.S. EPA, 1980). These guidelines were established to determine the concentration of toxicants in water that do not pose significant risks to the general population. Water quality criteria are derived for carcinogenic and noncarcinogenic effects. The guidelines for noncarcinogenic effects were based on selecting appropriate NOAELs or LOAELs and applying safety factors as deemed necessary to account for the uncertainty in using animal models as surrogates for toxic risk for humans. The water quality criteria guidelines evolved into the present day methodology for derivation of reference doses (RfDs) (U.S. EPA, 1993) and reference concentrations (RfCs) (U.S. EPA, 1994a) for assessment of noncarcinogenic systemic effects. The RfD/RfC is an estimate (with uncertainty spanning perhaps an order of magnitude) of exposure to the human population that is likely to be without appreciable risk of noncancer health effects during a lifetime (U.S. EPA, 1994a). The concept of threshold is inherent in the definition of the RfDs and RfDs; the concentration below which there is no observable adverse effect is considered to be a threshold (U.S. EPA, 1993). A threshold is assumed to exist for both the individual and the population (U.S. EPA, 1993). RfDs/RfCs are derived for chemicals that cause noncancer (or systemic) health effects. All effects except cancer and gene mutations comprise noncancer health effects; these include effects on portal-of-entry organs (gastrointestinal and respiratory tracts and skin), remote sites (internal organs such as liver, kidney, bone marrow, brain, reproductive organs), and the developing fetus. Risk assessment guidelines have been established for general noncancer endpoints (U.S. EPA, 1993, 1994a), development toxicity (U.S. EPA, 1991c), and neurotoxicity (U.S. EPA, 1994b). Guidelines were proposed for male (U.S. EPA, 1988a) and female reproductive toxicity (U.S. EPA, 1988b); the guidelines for male reproductive toxicity are under discussion. In addition to the above citations, EPA's RfC methodology has been discussed in several reports by Jarabek and coworkers (Jarabek et al., 1990, Jarabek, 1995a,b).

The general guidelines for deriving RfDs and RfCs are based on the risk assessment paradigm established by the NRC/NAS (1983): *hazard identification, dose-response assessment, exposure assessment*, and *risk characterization*.

4.1. Hazard Assessment

4.1.1. General Guidelines for RfD/RfC Derivation (U.S. EPA, 1989a, 1991b, 1993, 1994a)

Hazard identification concerns the careful evaluation of all relevant human and animal data and identifying the principal studies that best describe the statistically and biologically significant effect(s) expected to occur in the general population. The data sources for detailing systemic effects likely to occur in humans after exposure to environmental substances are the same as those identified for evaluating potential carcinogenicity. They include human studies (epidemiologic and case studies), long-term animal studies, short-term studies used to identify targets for long-term studies, toxicokinetics studies, and studies on mechanism of action and structure-activity relationships. For deriving RfDs, the evaluation focuses primarily on oral exposure studies and on inhalation studies for deriving RfCs.

EPA has established guidelines for assessing the quality of individual human and animal studies and for assessing the overall quality of the database (U.S. EPA, 1989a, 1991b, 1993, 1994a). Human studies are always given priority over animal studies for assessing the potential hazards of environmental exposures to the general population. Identification of sensitive populations is a critical aspect of hazard assessment. Epidemiologic and clinical studies conducted on potentially sensitive groups are helpful, as well as identification of genetic and other risk factors that may contribute to increased risk. Sensitive groups may include, groups in prenatal and postnatal developmental stages, those with respiratory diseases, circulatory conditions, and liver diseases, and the elderly.

When adequate human studies are not available, animal studies are used. Assessing the validity or appropriateness of animal models requires consideration of the study design (ideal studies are those that follow established procedures and protocols for conduct and analysis of results), elements of exposure definition (concentration, duration, frequency, route, etc.), relevance of exposure levels tested, similarities and differences between the test species and humans. The appropriate or most relevant species is identified based on comparable metabolism, pharmacokinetics, etc. The most sensitive species is not selected as a priority, because the effects produced may not be toxicologically relevant to human. If the most relevant species cannot be identified, then the most sensitive species is selected as a science policy option (U.S. EPA, 1994a). Route-to-route extrapolation of hazard concerns is permissible; EPA's view is that the toxicity potential manifested by one route can be indicative of potential toxicity via any other exposure route unless convincing evidence to the contrary exists.

The overall evaluation identifies the critical effect and the effect levels (NOAEL and LOAEL) associated with exposure to a substance. In addition, the overall conclusions (weight of evidence) regarding the likelihood of an environmental substance posing a hazard for humans is enhanced by the following factors (U.S. EPA, 1989a, 1991b): (1) a clear dose-response relationship, (2) similar effects across species, sex, strain, exposure routes, multiple experiments, (3) biological plausibility of the effect of concern, (4) similar effects in structurally related compounds, and (5) a link between the chemical and evidence of the effect of concern in humans. These factors also increase the confidence in the weight of evidence (U.S. EPA, 1989a). Other factors that increase the confidence in the weight of evidence concerns the completeness of the database. EPA has established criteria by which the completeness of the database can be judged when human data are not available. a complete database consists of (1) two long-term inhalation studies in different species, (2) a mammalian two-generation reproductive toxicity study, and (3) two mammalian developmental toxicity studies in different species. The minimal requirement for deriving an reference values is one long-term (preferred) or subchronic (acceptable) study. These factors also contribute to the confidence in the selection of the critical effect and effect levels.

4.1.2. Developmental Toxicity Guidelines

The U.S. EPA (1991c) defines developmental toxicology as "the study of adverse effects on the developing organism that may result from exposure prior to conception (either parent), during prenatal development, or postnatally to the time of sexual maturation". Developmental toxicity may be manifested as one or more endpoints: (1) death; (2) structural abnormality; (3) altered growth; and (4) functional deficit.

Data from all available sources, with primary emphasis on human and experimental animal data, are used to evaluate the developmental toxicity potential of a substance (U.S. EPA, 1991c). Human data are often inadequate for evaluating the potential of a chemical to cause developmental toxicity; therefore, it is often necessary to rely on experimental animal data. When using animal data for hazard identification of potential developmental toxicants, several assumptions are made: (a) a substance that causes developmental toxicity in animals will potentially pose a hazard to humans; (b) all four of the manifestations of developmental toxicity are of concern; (c) the types of developmental effects seen in animals are not necessarily the same as those that may occur in humans; and d) the most sensitive species is appropriate for use (U.S. EPA, 1991c).

Maternal toxicity endpoints considered for risk assessment include mortality, body weight and body weight change, food and water consumption, clinical evaluations, gestation length (if animals are allowed to deliver), mating and fertility indices, organ weights, and gross observations

(U.S. EPA, 1991c). Comparison of maternal toxicity in developmental toxicity studies with data from other toxicological studies establishes differences in the responses of pregnant animals to nonpregnant adults.

Developmental toxicity studies generally evaluate death, structural abnormality, and altered growth. Endpoints evaluated for risk assessment include implantation sites, corpora lutea, live and dead offspring, resorptions, pre- and postimplantation loss, altered offspring (external, visceral or skeletal malformation or variations), sex ratio, and fetal body weight (U.S. EPA, 1991c). Testing for functional deficits, which may occur between conception and sexual maturation, is not routinely required. Most of the work involving functional evaluation has been in the area of developmental neurotoxicity and testing protocols and data interpretation are beginning to be standardized (Francis, 1992). Analyses of postnatal renal development have detected what has been interpreted as either apparent hydronephrosis (a variation) (Woo and Hoar, 1972) or hydronephrosis (a malformation) (U.S. EPA, 1991c). Other systems that are not as well studied include the cardiovascular, respiratory, immune, endocrine, reproductive, and digestive systems (U.S. EPA, 1991c).

Dose levels of a substance that results in maternal toxicity may differ from those that result in developmental toxicity. Of greatest concern, are those agents that result in developmental toxicity with no apparent maternally toxic effects. For most substances however, the exposure situations of concern are those potentially near the maternally toxic dose level, but developmental effects should not be considered secondary to maternal toxicity. At dose levels of a substance that result in marked maternal toxicity, stress associated with disruption of homeostasis can occur. Fetal anomalies commonly linked to maternal toxicity include bent or wavy ribs, reduced weight, and death (Black and Marks, 1992; Khera, 1984). However, such fetal effects are toxic manifestations and are considered in regulation and risk assessment since maternal effects may be reversible, whereas effects on the fetus may be permanent (U.S. EPA 1991c).

Substances are categorized based on evidence for developmental toxicity. *Sufficient human evidence* includes data from epidemiological studies that provide strong scientific evidence for a causal relationship. *Sufficient experimental animal evidence/limited human data* includes animal studies and/or limited human data that provide convincing evidence of the potential for developmental toxicity. The *insufficient evidence* category includes substances for which there is limited data upon which to base a scientific judgement, because of no studies or inadequate studies.

4.1.3. Neurotoxicity Guidelines

Hazard identification for neurotoxicity evaluates relevant data to determine if a substance is likely to cause an adverse effect in the nervous system. As for the RfD and RfC, hazard identification for neurotoxicity may involve evaluation of data from various sources including epidemiologic studies, clinical evaluations in humans, short- and long-term animal studies, mechanism-of-action studies, and structure-activity relationship studies (U.S. EPA, 1994b). Generally, hazard identification will evaluate mechanism of action data as well as human and animal toxicity data. Hazard identification for neurotoxicants is complicated by the difficulty in specifically defining an adverse neurological effect, thereby resulting in widely varying estimates of the number of neurotoxic chemicals. Tilson (1990) noted that neurotoxicants are those agents that adversely affect the neurophysiological, neurochemical or structural integrity of the nervous system or the integration of nervous system function expressed as modified behavior.

According to U.S. EPA risk assessment guidelines for neurotoxicity (U.S. EPA, 1994b), direct identification of hazard may be obtained from human studies but human data are often anecdotal, involve acute exposures causing overtly toxic or lethal effects, or (especially for epidemiologic studies) are often complicated by confounding factors such exposures to multiple agents or imprecise exposure characterization. For human studies, identification of neurotoxicity may include evaluations of neuromuscular strength, alterations of sensory-motor function, learning and memory deficits, personality and mood alterations, and alterations in autonomic functions.

Animal studies using models that measure behavioral, neurophysiological, neurochemical, or structural changes are available and can be used to extrapolate to humans for the purpose of hazard assessment. With the exception of behavioral changes, it is generally assumed that neurological processes are fundamentally similar among most species and, therefore, the effects observed in one species are likely to occur in another species including humans. Uncertainties (e.g., sex-dependent differences, species specificity, etc.) are inherent in such extrapolations, however, and special issues often arise regarding how these differences impact on hazard identification in humans.

Assessing the validity and appropriateness of human and animal studies for neurotoxicity hazard identification is generally similar to the methods and procedures described for the RfD/RfC. However, concern has been expressed regarding the fact that identification and acceptance of a NOAEL from a single study may not necessarily be indicative of an absence of neurotoxicity risk.

The objectives of neurotoxicity testing include: (1) determining if the nervous system is affected by the toxicant (detection), (2) characterizing alterations of the nervous system that are associated with exposure to the toxicant, (3) ascertaining if the nervous system is the primary target, and (4) determining the dose-effect and time-effect relationships relative to establishing a no-observable-effect-level (NOEL) (Reiter, 1987).

4.2. Dose-Response Assessment

4.2.1. Guidelines for NOAEL/LOAEL Approach (U.S. EPA, 1989a, 1991b, 1993, 1994a)

4.2.1.1. General Methodology – A dose-response relationship is seen when increased dosage of a toxicant results in increases in a response to the toxicant; the response may be in the form of an increased incidence (quantal response), increased severity (graded response), or increased incidence and severity of an effect. Dose-response assessment is dependent initially on hazard assessment activities: selecting the principal study and selecting the critical effect. In selecting the principal study, human studies are considered first if quantitative exposure and incidence or severity information can be obtained from these studies, because human data eliminate the need for species extrapolation. If a suitable human study is not available for quantitative assessment, then the principal study is selected from the available animal studies. The study using an animal model (species) most relevant to humans (based on mechanism of action, pharmacokinetics, route of exposure, etc.) is given first priority. If no relevant animal model can be identified, then the most sensitive animal species is chosen (default position), because there is no assurance that humans are not as sensitive as the most sensitive species. The most sensitive species is one showing a toxic effect at the lowest tested dose. Therefore, the critical effect is the effect occurring at the lowest dose in either the most relevant or most sensitive species; the NOAEL corresponding to the critical effect is selected for deriving the RfD or RfC. The dose or exposure concentration may be measured as the applied dose or concentration, absorbed dose, or target organ dose. It should be noted that, for portal-of-entry effects, the absorbed dose may not be a relevant expression of exposure. If the critical effect is prevented then all other toxic effects should be prevented (science policy). Exposures less than the reference value are considered to be without (but not categorically) risk of adverse effects in humans.

The RfD or RfC is calculated by the following simple equation:

$$\text{RfD or RfC} = \text{NOAEL}/(\text{UF} \times \text{MF}),$$

where UF is the uncertainty factor and MF is the modifying factor. The RfD or RfC can be derived from a LOAEL when an NOAEL cannot be identified from the principal study. The RfD is expressed in mg/kg/day and the RfC in mg/m^3/day.

Dose scaling for species extrapolation in quantitative assessment for noncancer toxicants does not take the same form as for carcinogens. Uncertainty factors are used in noncancer assessments to account for the pharmacokinetic and pharmacodynamic differences between humans and animals, in contrast to the dose scaling procedures (body weight to ⅔ or ¾ power) used for cancer risk assessments.

For the inhalation exposures, human equivalent concentrations (HEC) are calculated according to the following steps: (1) conversion of exposure units from ppm to mg/m^3, (2) adjustment of experimental exposure to 24-h continuous exposure for a lifetime of 70 years for humans, and (3) adjustment of doses for the type of substance (particle/aerosol or gas/vapor) and the anatomical site of the effect (respiratory or extrarespiratory). Pharmacokinetics data are be used for dosimetric adjustments when available. Physiological parameters including surface areas for the different regions of the respiratory tract for humans and experimental animals, body weight, and ventilatory values are used to calculate the HEC. The HEC is calculated according to the following equation:

$$NOAEL_{[HEC]}\ (mg/m^3) = NOAEL_{[ADJ]}\ (mg/m^3) \times DAF_r,$$

where $NOAEL_{[ADJ]}$ is the effect level adjusted for discontinuous exposure and DAF_r is the dosimetric adjustment factor for respiratory region of concern based on the regional deposited dose ratio (RDDR) for particles or the regional gas dose ratio (RGDR) for gases. The RDDR or RGDR is the ratio of the regional deposited dose or regional gas dose (RDD_A or RGD_A) to a target in animals and the regional deposited dose or regional gas dose (RDD_H or RGD_H) to a target in humans. These ratios adjust the exposure concentrations used in animal studies to estimate the dose delivered to a target region in humans. A detailed discussion on the methods for calculating regional doses is found in EPA's reference concentration methodology (U.S. EPA, 1994a).

Dosimetric adjustments for particles and vapors takes into account particle size and size distribution of the particles. Dosimetric adjustments for gases is determined by the physicochemical and uptake characteristics of the gas. Category 1 gases are highly water soluble, rapidly irreversibly reactive, and does not accumulate in blood. Category 2 gases are moderately water soluble, rapidly reversibly reactive, and may accumulate in blood. Category 3 gases are

insoluble in water, unreactive in the extrathoracic and tracheobronchial regions, and their site of action are remote to the respiratory tract (extrarespiratory region or systemic). Default equations for estimating doses to the respiratory region of concern have been derived for both particles and gases (U.S. EPA, 1994a). These equations can be used when chemical specific data are not available for dosimetric adjustments. More details about the regional deposition of particles or regional gas effect/uptake of substances in the respiratory tract and the ratios between humans and animals for calculating human equivalent concentrations are presented in EPA's guidelines (U.S. EPA, 1989a, 1994a) and other reports (Jarabek et al., 1990, Jarabek, 1995a,b; Shoaf, 1991).

Uncertainty factors are applied to the $NOAEL_{[HEC]}$ to derive the RfD/RfC. The uncertainty factors account for the following: (2) variations in sensitivity among the human population, (2) extrapolating from animal data to human exposures, (3) less than lifetime exposures when subchronic studies are used instead of chronic studies, (4) extrapolating from a LOAEL to a NOAEL when a suitable NOAEL cannot be identified, and (5) an incomplete database. Uncertainty factors of 10 are usually applied to derivation of RfDs. For deriving RfC, uncertainty factors of 10 are applied for variation in human sensitivity, up to 10 for a subchronic study, LOAEL, and incomplete database, and 3 for animal to human extrapolation. In addition, a modifying factor of 10 or less can be applied to account for scientific uncertainties in the study or database not accounted for by uncertainty factors.

4.2.1.2. Approach Used for Developmental Toxicity – If developmental toxicity occurs as a result of exposure a substance, developmental endpoints are evaluated quantitatively using a protocol similar to those for deriving RfDs or RfC. An oral or dermal reference dose (RfD_{DT}) or an inhalation reference concentration (RfC_{DT}) for developmental toxicity is derived (U.S. EPA, 1991c). The RfD_{DT} or RfC_{DT} is derived by dividing the NOAEL or LOAEL by the total uncertainty factor. NOAELs or LOAELs are determined from the most sensitive, or critical, developmental effect from the most sensitive animal species. If the NOAEL or LOAEL for maternal toxicity is lower than that for developmental toxicity, this should be noted in the risk characterization and the value compared with data from other experiments of adult exposure. Uncertainty factors generally include a 10-fold factor for interspecies variation and a 10-fold factor for intraspecies variation (3 for RfC) ; an uncertainty factor is not applied for duration of exposure.

The dose-response relationship is evaluated in standard animal studies using three dose groups and a control. For developmental toxicity studies, a threshold is assumed for the dose-response relationship (U.S. EPA, 1991c). Three general patterns of response have been described for agents that cause developmental toxicity (Manson, 1986). Substances that have high

developmental toxicity potency can cause malformations of the entire litter at dose levels that do not cause embryolethality. Conversely, some substances may result in growth retardation and death without malformations. The more common dose-response pattern includes some embryolethality with growth retardation and malformations evident in surviving fetuses. Generally, as dose levels increase causing embryolethality to increase, an observed decrease in malformations may result.

4.2.1.3. Approach Used for Neurotoxicity – If chronic or subchronic toxicity studies show evidence of neurotoxicity and batteries of neurotoxicity test are conducted, it may be possible to derive an RfD or RfC for neurotoxicity. The dose-response assessment is conducted as using the RfC/RfD approach as described above.

Alternately, the benchmark dose concept has been suggested because it makes use of the dose-response curve rather than single dose (i.e., NOAEL) for estimating potential risk. Rather than extrapolating to doses far below the experimental dose range, the benchmark dose estimates a the dose corresponding to a specific incidence of an effect (e.g., 10%) based on the upper confidence limit of the particular dose-response curve (Farland and Dourson, 1993).

4.2.2. Benchmark Approach (U.S. EPA, 1995; Barnes et al., 1995)

The benchmark dose (BMD) has been defined as the statistical lower confidence limit for a dose that produces a predetermined change in response rate of an adverse effect (benchmark response or BMR) compared with background (Crump, 1984a). The benchmark approach can be applied to general noncancer dose-response data, as well as developmental, reproductive, and neurotoxicity data. It has many features in common with deriving the LED_{10} for cancer data. The BMD is calculated using dose-response data fitted to mathematical curve-fitting models with appropriate statistical procedures. There is no extrapolation to doses below the experimental range, but to a predetermine BMR of 1, 5, or 10%.

The EPA is developing guidelines for applying the BMD approach to noncancer dose-response assessments, but the guidelines are not presently available. A overview and description of the benchmark approach has been presented by EPA's Risk Assessment Forum (U.S. EPA, 1995), and a workshop organized to discuss the feasibility and implications of the BMD approach for derivation of reference values (Barnes et al., 1995).

There were several reasons for developing an alternative approach for deriving reference values for noncancer health effects. The primary reason deals with the limitations of the

NOAEL/LOAEL approach for RfD/RfC derivation. These limitations include the following (Crump, 1984a; U.S. EPA, 1995; Barnes et al., 1995):

- The NOAEL is based on scientific judgment and may be the source of controversy.
- Experiments involving a few animals tend to produce large NOAELs, and consequently, large RfDs/RfCs.
- The slope of the dose response plays little role in determining the NOAEL.
- The NOAEL is limited to the experimental doses tested and is dependent on the statistical power of the study.

The benchmark approach to deriving an RfD/RfC involves three steps that are discussed below:

(1) Selection of the response or group of responses from the experimental data set. The responses from animal studies are selected based on toxicological relevance to humans and convincing evidence of a dose-related effect for the responses. Mathematical curve-fitting should be applied to all relevant responses.

(2) Calculation of the BMDs for the selected responses. Calculation of the BMD may involve transforming the data to a form that can be fitted by a mathematical model. Fitting data presented in the quantal format (dependent only on presence or absence of a response), such as incidence data, is a straightforward process using most mathematical models (Barnes et al., 1995). Categorical data in which effects are described in terms of severity of effect (mild, moderate, severe, etc.) and continuous data (e.g. body and organ weights, enzyme levels, etc.), which can be transformed into a quantal format or modeled without being transformed is a more complex process (Barnes et al., 1995). Sometimes data are presented in more than one format, i.e., the severity of liver lesions may presented along with the incidence of the lesion. The mathematical model chosen to estimate the BMD depends on the format of the data. For quantal data, the probability of a response or the dose (BMD) corresponding to a specific response (BMR) can be estimated. For continuous data, the mean response corresponding to dose can be estimated. Different types of mathematical models, such as quantal, quadratic, and polynomial regression can be fitted to both quantal and continuous data; Weibull and log-normal can be fitted to quantal data and linear-quadratic and continuous models can be fitted to continuous data (U.S. EPA, 1995). The model selected should adequately describe the biological response, such as the slope transitions near the threshold. After the model is selected, the BMR level is selected; the value is generally in the range of 1 to 10%.

(3) Choosing an appropriate BMD and calculating the RfD/RfC. Because multiple studies with multiple responses may be available or a single response may be subjected to multiple curve-fitting models, multiple BMDs can sometimes be calculated for a single assessment. Three options that have been recommended for selecting the BMD are as follows: (1) calculate an average or geometric mean of the BMDs, (2) use the most appropriate species and/or sex, or (3) select the smallest BMD. After the BMD is selected, the RfD/RfC is calculated by applying uncertainty factors. Several options were presented for choosing the uncertainty factors: (1) use uncertainty factors similar to those applied to NOAELs, (2) use uncertainty factors applied to NOAELs modified by the average ratio of BMD/NOAEL, (3) use-risk based uncertainty factors to extrapolate to 10^{-4} or 10^{-5}, which would represent a virtually safe dose (Kimmel and Gaylor, 1988), (4) use uncertainty factors dependent on the choice of BMR and size of confidence limit, and (5) use uncertainty factors that incorporate the slope of the dose-response and/or other biological considerations.

5. ISSUES IN HUMAN HEALTH RISK ASSESSMENT

This report has focused on the EPA's cancer and noncancer risk assessment guidelines and methodologies. There are similarities and differences in the assessment procedures used for the two types of hazards. The benchmark approach for noncancer assessments has been harmonized with the dose-response analysis within the range of experimental observation for carcinogen assessments; the methodology for dosimetric adjustments for inhaled particles and gases are the same for cancer and noncancer hazards; and the criteria for assessing the adequacy of human and animal data are similar. Some of the issues can be applied to both cancer and noncancer assessments: route-to-route extrapolation, pharmacokinetic modeling, target organ concordance involving non-analogous sites, and including the sensitive population in the LED_{10}. Nevertheless, the following issues are discussed as they relate to either the cancer and noncancer risk assessments. The issues are discussed in detail in the following sections, and some key issues are summarized briefly in Section 6.

5.1. Carcinogen Risk Assessment

The following sections address some of the issues concerning the cancer risk assessment process. Some of these issues were specifically discussed in the EPA's1996 proposed cancer guidelines in its discussion on default assumptions (U.S. EPA, 1996b).

5.1.1. Relevance of Animal Models to Human Carcinogenicity

EPA's 1986 guidelines state that each animal study should be reviewed as to the relevance of the evidence for humans. In a working paper (draft), EPA stated that ".....tumors at any animal tissue site support an inference that humans may respond at some site" (science policy) (U.S. EPA, 1992a). The agency further stated in the same report that if the information suggests a mechanism unique to an animal species or strain, then the evidence does not support a carcinogenic hazard to humans. This concept was slightly modified in the 1996 guidelines, which stated that "Information on an agent's potential mode(s) of action is important in considering the relevance of animal effects to assessment of human hazard."

Munro (1988) studied the relevancy of animal models to human carcinogenicity and proposed eight criteria by which relevancy could be judged. Evidence would not be considered as relevant to human carcinogenicity if these criteria can be applied to the data gathered on a particular substance (Munro, 1988).

1. The test chemical or mixture does not represent that to which humans are exposed.

2. The route of exposure of the test animal is vastly different from that of humans (route-to-route extrapolation).

3. The only tumor response of the test species occurs at a site having high background incidence (high spontaneous incidence).

4. The tumor response occurs only at high doses that produced toxicity incompatible with normal physiological function (carcinogenicity only at high doses).

5. Tumors are produced only at an anatomical site not found in humans (target organ concordance).

6. Tumor induction is closely linked to chronic physical irritation, physiological perturbations, or a marked derangement of endogenous metabolism.

7. Pharmacokinetic studies show vast differences in the disposition or fate of the test material between the animal model and humans or between the low and high doses.

8. Epidemiologic evidence suggests that the substance is not carcinogenic in humans under normal conditions of exposure.

Most of these items were addressed in EPA's 1996 proposed cancer assessment guidelines. However, EPA's guidelines and Munro's criteria differ in some respects. According to the guidelines the only criteria for judging whether hazard can be extrapolated to a second route is absorption of the agent by that route to give an internal dose. The 1996 guidelines further state that adequate data are necessary to demonstrate that an agent will act differently by a second route. Therefore, the burden of proof is on showing that an agent will act differently rather than acting similarly by a different route. First pass effects, particularly those occurring as a result of liver alterations after absorption from the gastrointestinal tract, are not taken into account. The default assumption allowing route-to-route extrapolation for hazard assessment is very conservative. Quantitative route-to-route extrapolation is conducted on a case-by-case basis. In most cases, however, sufficient data will not be available to conduct quantitative route-to-route extrapolation. Therefore, the issue is: should qualitative route-to-route extrapolation be allowed when sufficient data are not available for quantitative extrapolation?

Target organ concordance a carcinogen assessment issue, particularly for organs or tissues not shared by animals and humans. EPA's view is that target organ concordance is not assumed *a priori* for establishing potential carcinogenicity for humans (U.S. EPA, 1996b). Gregory (1988) stated that site-specific tumors should not be unique to the species tested and should be correlated with carcinogenicity in humans. According to Goodman and Wilson (1991), site concordance is more likely to occur if the routes of exposure of the animal models and humans are similar. Munro (1988) listed anatomical sites not shared by test species and humans as one of the criteria for

judging the evidence as not relevant and gave as an example the production of forestomach tumors in rats fed butylated hydroxyanisole (BHA). Munro (1988) also considered other criteria (high doses, irritative hyperplasia, and lack of genetic toxicity) in questioning the relevancy of this evidence for potential carcinogenicity of BHA in humans. Questions related to this issue are: Should target organ concordance be considered in establishing relevancy of animal models for human cancer assessments? Should organs unique to animals, such as the forestomach and Zymbal's gland be given special consideration? Should target organ concordance be considered when carcinogenicity is seen only in organs such as the forestomach? Does forestomach tumors occurring after oral exposure suggests that exposure by another route would result in tumor development in the contact organs? For example, would skin tumors develop after dermal route of exposure or the respiratory tract tumors after inhalation exposure? Although the default assumption described in the 1996 proposed cancer guidelines states that target organ concordance is not a prerequisite for establishing potential carcinogenicity to humans, carcinogenicity occurring only in target organs unique to animals poses a problem for weighing evidence of carcinogenicity.

In 1986, EPA's position regarding high background tumors was that an increased incidence of tumors with a high background constituted "sufficient" evidence of potential carcinogenicity, but when other evidence (replicate studies, malignancy) is considered the conclusion may be changed (U.S. EPA, 1986). According to the 1996 proposed guidelines, tumors with high background incidences are considered in a hazard assessment, but are given less weight (U.S. EPA, 1996b). Liver tumors in male mice, testicular interstitial cell tumors in male rats, and pituitary tumors in male and female rats occur with high background rates (Gregory, 1988; Goodman and Wilson, 1991). Mouse liver tumors are often used as evidence of potential carcinogenicity in humans. The background incidence of liver tumors the $B6C3F_1$ male mouse was reported to be as high as 31.1% (Pereira, 1985); the corresponding incidence in females was only 6.2%. A different pattern of activated oncogenes is seen in spontaneous liver tumors compared with those induced by chemicals (Reynolds et al., 1988). Therefore, in cases where the background tumor incidence is high, it may be possible to distinguish between increased incidences due to stimulation of existing lesions and the induction of new tumors (Anonymous, 1993).

Another issue related to relevance of animal models to human cancer risk is carcinogenesis by a mode of action known not to occur or unlikely to occur in humans. EPA's position is that all mode of action information should be considered when assessing risk, and if the mode of action has been shown not to occur in humans, the agent is *not likely* to be carcinogenic to humans. One example is tumorigenesis in male rat kidney due to accumulation of $\alpha_{2\mu}$-globulin, a mode of action

that does not occur in humans. EPA has established a science policy and guidance for assessing the carcinogenicity of chemicals inducing renal tubule tumors (U.S. EPA, 1991a). The science policy is as follows: *Male rat renal tubule tumors arising as a result of a process involving $\alpha_{2\mu}$-globulin accumulation do not contribute to the qualitative weight-of-evidence that a chemical poses a human carcinogenic hazard. Such tumors are not included in dose-response extrapolations for the estimation of human carcinogenic risk.* The guidance state that the assessor first evaluate renal tubule tumor data to determine if $\alpha_{2\mu}$-globulin is involved in the carcinogenic process. If so, then the extent to which $\alpha_{2\mu}$-globulin accounts for the carcinogenicity as opposed to other mechanistic processes is evaluated (U.S. EPA, 1991a).

In addition, urinary bladder tumors associated with formation of calculi or implantation of foreign bodies into the bladder lumen should be evaluated carefully as well as tumors formed after exposure by routes not likely to be encountered by humans (intraperitoneal or subcutaneous injection).

The susceptibility of rats and mice to urinary bladder carcinogenesis induced by endogenous formation of a bladder calculus or a foreign body implanted into the bladder lumen has been documented (Cohen and Ellwein, 1992). Chemicals producing calculi are generally nongenotoxic, induce tumors above a threshold (Cohen and Ellwein, 1992), and involve a mode of action not likely to occur in humans. Consequently, such tumors are not likely be relevant to evaluating potential carcinogenicity to humans. Bladder tumors induced by implantation of substances into the bladder lumen should also be considered as not relevant to human carcinogenicity. OSTP (1985) also stated that such tumors may not be relevant to human oral exposure.

There are some routes of exposure by which humans are unlikely to be exposed, such as subcutaneous, intraperitoneal, intratracheal instillation. OSTP (1985) noted that subcutaneous sarcomas formed by subcutaneous injection of a substance may not be relevant to human exposure. How much weight should be given to evidence of carcinogenicity by unusual routes of exposure when evaluating potential carcinogenicity, particularly in cases where local tumors are formed? Should evidence from a route unlikely to be experienced by humans be considered in a different light from evidence of carcinogenicity induced in animals at high doses not likely to be experienced by humans?

5.1.2. Genotoxic vs Nongenotoxic Modes of Action

Distinguishing between genotoxic and nongenotoxic modes of action in tumor induction is a critical element in assessing the potential carcinogenicity of agents to humans. Purchase (1994) defined genotoxic carcinogens as those chemicals or their metabolites that alter DNA or genetic information by producing point mutations, insertions, deletions, or changes in chromosome structure or number. The outcome of evaluating the genotoxicity of a substance is to determine if the substance interacts directly with DNA resulting in damage to DNA or heritable changes in the DNA (Cohen and Ellwein, 1992). The DNA damage could result in activation of oncogenes (e.g. *ras* or *myc*) or alterations of tumor suppressor genes (Purchase, 1994). Carcinogenicity induced by genotoxic substances is considered to have no threshold, i.e., that every exposure no matter how small is associated with a measurable risk. Most chemicals known to be carcinogenic to both humans and animals have genotoxic activity (Cohen and Ellwein, 1992).

Nongenotoxic carcinogens are described as those chemicals whose primary activity does not involve genetic toxicity. According to Purchase (1994), nongenotoxic carcinogens induce mitogenesis and hyperplasia which allows the fixation of DNA damage by oxygen free radicals or different types of mutational events produced by other endogenous or exogenous agents. There are several mechanisms by which nongenotoxic agents induce neoplasia (Purchase, 1994). Some of the more widely studied are (1) excess production of trophic hormone or disruption of homeostasis (thyroid stimulating hormone/thyroid neoplasia; trypsin inhibitors/pancreatic neoplasia; gastrin/ECL gastric neoplasia; luteinizing hormone/Leydig cell neoplasia), (2) receptor binding (dioxin:Ah cytoplasmic binding), and (3) cytotoxicity and mitogenesis (crystalline formations/bladder neoplasia, $\alpha_{2\mu}$-globulin/male rat kidney neoplasia). Purchase (1994) noted that nongenotoxic carcinogens usually display tissue (affect a single tissue or organ) and species specificity (affect only one species or one sex in one species).

Potential carcinogenicity in humans is often assessed based on induction of mouse liver tumors. Pereira (1985) stated that a distinction should be made between genotoxic and nongenotoxic modes of action when assessing potential carcinogenicity to humans based on mouse liver tumor data. The nongenotoxic carcinogens may act by promoting already initiated cells in the mouse liver, and therefore, may promote spontaneously or environmentally initiated cells in human liver. He further stated that the anticipated nonlinear dose-response curve for nongenotoxic carcinogens does not mean that these chemicals pose no risk to humans, but that a safe level may exist. Reynolds et al. (1988) reported that spontaneously occurring mouse liver tumors and those induced by chemical agents have different patterns of activated oncogenes. Different patterns

were also seen in rat tumors that developed spontaneously compared with lung tumors induced by chemical agents.

EPA's 1986 guidelines stated that information on genotoxicity provides supportive evidence of carcinogenicity and may indicate the mode of action. In the 1996 proposed guidelines, the genotoxic nature of an agent is evaluated within the concept of mode of action, which is an integral part of the cancer risk assessment process. The 1996 proposed guidelines also noted that mutagenic chemicals usually induce tumors across species and at multiple sites, with both situations increasing the level of concern about a chemical's carcinogenic potential in humans. Issues of evaluating nongenotoxic substances for potential carcinogenicity in humans are interrelated with issues of cell proliferation and mode of action. The technical review workshop reviewer who evaluated EPA's proposed guidelines listed three types of genotoxic effects as subsets of mode of action: direct mutagenic effects, indirect mutagenic effect, and heritable epigenetic effects (Eastern Research Group, 1994).

To establish the genotoxic and nongenotoxic modes of action, it is necessary to develop criteria for judging the genotoxicity of an agents and determine how much evidence is sufficient to determine that an agent is genotoxic.

5.1.3. Carcinogenicity as a Manifestation of Cell Proliferation or Toxicity

Huff (1993) analyzed site-specific cell proliferation/toxicity and carcinogenic responses in long-term toxicity studies using rodents. He evaluated 53 chemicals tested in male and female Fisher rats and B6C3F_1 mice producing a total of 207 carcinogenic responses. He examined proliferative, toxic, and carcinogenic responses to specific chemicals (1,4-dichlorobenzene and furan), and evaluated specific target organs (liver and kidney). Huff (1993) concluded that "toxicity findings from higher exposures in short-term experiments or observed toxicity in long-term experiments (e.g., kidney) cannot be used mechanistically to either predict eventual carcinogenicity or to advocate toxicity and resultant sequelae as a mechanism of tumor development." He, therefore, questioned the notion of using cellular proliferation and toxicity measured for 1 week as a means for establishing the mechanism for liver carcinogenesis; he also noted that the species and sex showing liver toxicity and hepatocyte proliferation may not be the one showing a carcinogenic response. Huff (1993) also noted that spontaneous chronic nephropathy could not be correlated with renal carcinogenesis in rats. The high incidence or severity of nephropathy is not necessarily associated with a high incidence of renal cancer. From his evaluation, Huff (1993) reached six major conclusions concerning toxicity and carcinogenicity in long-term rodent studies:

1. Only 7 of the 53 carcinogenic chemicals produced target organ toxicity at all carcinogenic sites.

2. Only three chemicals showed carcinogenicity at the highest dose without corroborating evidence at lower doses (refutes the "high-dose-only carcinogen" theory).

3. The number of chemicals with a possible "indirect or secondary mechanism" (i.e,, toxicity) is small.

4. There is no uniform correlation between induction of toxicity and carcinogenicity.

5. Chemicals evaluated for long-term toxicity and carcinogenicity fall into three categories:
   a. those causing organ toxicity without cancer,
   b. those causing cancer without associated target organ toxicity,
   c. those causing both site-specific toxicity and carcinogenicity.

6. To separate chemicals by mechanisms of carcinogenicity (e.g., primary and secondary) for risk assessment purposes is premature

According to Huff (1993), it "would be premature and probably incorrect to make public health decisions on the basis of skimpy scientific data regarding the influence of cell proliferation *per se* on the carcinogenesis process." He further stated that cell proliferation had an influence on carcinogenesis, but existing data do not support the hypothesis that increased cell proliferation leads to or causes cancer, and there are no data suggesting that a noncarcinogenic chemical can be made carcinogenic by enhancing cell proliferation of a normal tissue. He asserted that more and better data are needed to establish the relationship between cell proliferation, toxicity, and carcinogenicity.

Ward et al. (1993) also looked at the correlation between toxicity, cell proliferation and carcinogenicity at specific sites (nasal cavity, liver, kidney, skin, and urinary bladder) and showed many instances where toxicity and cell proliferation did not lead to carcinogenesis. Ward et al (1993) suggested several reason for the lack of correlation: (1) cell proliferation does not occur in the stem cell population, which is an important target for carcinogens; (2) the sustained cell proliferation occurs before preneoplastic cells or foci appear, then the effects of cell proliferation are minimized; (3) cell proliferation may not play a role in some specific cases or in all cases in which it occurs.

Haseman (1985) also noted that, when considering the relationship between cytotoxicity and carcinogenicity, tissue damage does not always lead to carcinogenicity. Griesemer (1992) noted the lack of correlation between sites of early or late toxic effects and carcinogenesis in animal studies.

In contrast, Moolgavkar (1993) presented two reasons for a role of cell proliferation in carcinogenesis. First, the increase in cell proliferation leads to an increase in mutation frequency and consequently an increase in the risk of cancer. Second, the increase in cell proliferation relative to cell differentiation or cell death results in a larger population of cells susceptible to malignant transformation. He concluded that programmed cell death and cell proliferation are important determinants of cancer risk.

Cohen and Ellwein (1993) stated that the bladder carcinogen *N*-[4-(5-nitro-2-furyl)-2-thiazolyl]formamide (FANFT) and the bladder and liver carcinogen 2-acetylaminofluorene (AAF) induce cell proliferation in the bladder epithelium at the high doses and are genotoxic after metabolic conversion to a reactive intermediate producing adducts. Carcinogenicity is not detected at doses below that which induces cell proliferation, although with AAF, DNA adducts were detected in the bladder epithelium at very low doses. In the liver, however, carcinogenicity occurred at doses below those associated with cell proliferation. Cohen and Ellwein (1993) concluded that a carcinogenic response was not observed in the bladder at the low genotoxic doses because of the detection power of the bioassay was limited; therefore, a true threshold was not established. In contrast, tumors induced by implantation of a foreign substance or the formation of calculi in the bladder lumen appear to act by a true threshold mechanism associated with the induction of cell proliferation. How can we be sure that detection limit of a bioassay is not a factor in observing a carcinogenic response with other agents considered to have a threshold?

The nasal cavity is another site in which carcinogenesis has been linked to cell proliferation. Monticello et al. (1993) reported that regenerative cell proliferation in the nasal cavity is clearly involved in carcinogenic response to formaldehyde, but it is not the only determinant of nasal carcinogenesis. Other toxicants, such as dimethylamine, that induce cell toxicity, inflammation, and squamous metaplasia, do not induce nasal tumors (Monticello et al., 1993). Additional data are needed to show the relationship between toxicity, epithelial proliferation, and carcinogenesis in the nasal cavity. A simple measurement of cell proliferation is inadequate for determining whether exposure to a chemical may be associated with increased cancer incidence (Short, 1993).

Rapid proliferation of cells has been postulated to increase the probability of a mutation occurring in the cells. Although this hypothesis is widely accepted as a mode of action for carcinogenesis, there is no evidence for an increased frequency of mutations in rapidly proliferating cell populations. Since no mutations have been detected as a result of cell proliferation, what is the role of cell proliferation in carcinogenesis?

5.1.4. Weight-of-Evidence Classification or Hazard Description

Hazard assessment usually concludes with a categorization of the evidence based on the degree to which the risk assessor believes the data support a causal association between cancer and exposure to a substance. Chemical classification has been the subject of concern, particularly the issues of whether chemicals should be classified, at what stage of the risk assessment they should be classified, and the basis for the classification. Anderson et al. (1993) stated that hazard conclusions should be delayed until the hazard assessment, dose-response assessment, and exposure assessment are brought together for an integrated summary. Scientists have questioned the "carcinogen" label placed on substances, because of the societal impact of this label (Harvard Center for Risk Analysis, 1994)

In EPA's 1986 guidelines, the classification of potential carcinogens was based on the a, B, C scheme similar to that used by IARC. This scheme is based on a three step approach, (1) determining the level of evidence in humans and animals (2) combining the level of evidence in humans and animals for a tentative categorization, and (3) applying supporting evidence to determine if the tentative categorization should be changed.

In one of its working paper (U.S. EPA, 1992a), EPA proposed a classification scheme whereby human and animal evidence of carcinogenicity would be classified (or weighed) separately. Briefly the categories for human evidence were described as follows: category 1- plausible evidence; category 2 - suggestive evidence; category 3 - inconclusive evidence; category 4 - evidence of noncarcinogenicity. Similar descriptive categories were proposed for animal evidence. Categorizing the animal evidence included an evaluation of the supporting evidence (short-term studies, genotoxicity data, structure/activity relationships, etc.) . The working group noted that no weight should be given to animal evidence not considered relevant to establishing potential human carcinogenicity. Therefore, relevancy would be included in the initial evaluation of the data instead of later in the classification. The final step in hazard characterization proposed by the working group was an overall route-specific weight-of-evidence scheme that would combine the categories for human and animal evidence. The hazard descriptors for the overall weight of evidence would be "*known*" (category 1 human evidence); "*highly likely*" (category 2 human evidence plus category 1 animal evidence, or strong category 1 evidence for animals, or "known" by one route and absorbed by another route); "*likely*" (category 1 animal evidence that is persuasive or category 2 for both human and animal evidence); "*some evidence*" (category 2 human or animal evidence); "*not likely*" (category 4 human or animal evidence). The hazard descriptor would be accompanied by a short hazard narrative that would characterize the evidence, discuss mechanism of action, and suggest an approach for the dose-response assessment. This

hazard classification scheme is rather confusing and has been replaced by the categories described in the 1996 proposed guidelines.

The NRC/NAS (1994) Committee recommended establishing four categories based on strength of evidence and relevance to humans: Category I - the substance might pose a carcinogenic hazard to humans under any condition of exposure; Category II - the substance might pose a carcinogenic hazard to humans, but under limited conditions of exposure; Category III - not likely to pose a carcinogenic hazard to humans under any conditions; Category IV - available evidence demonstrate a lack of carcinogenicity or no evidence available. This last category Implies that "no data" situations would be equivalent to evidence of noncarcinogenicity. The technical review workshop (Eastern Research Group, 1994) recommended that EPA adopt the scheme proposed by the NRC Committee.

Ashby et al. (1990) proposed a classification scheme consisting of eight categories. The classification considered human studies, animal bioassays, supportive evidence from bioassays, mechanistic studies for establishing potential human cancer hazards. Category 1: *"known human carcinogen"* - sufficient evidence for human carcinogenicity (a causal relationship is demonstrated between exposure to an agent and human cancer). Category 2: "*carcinogenic activity in animals; probable human carcinogen*" - limited evidence of carcinogenicity in humans (a causal relationship is observed, but biases and confounders cannot be ruled out) and sufficient or limited evidence in animals along with evidence showing the carcinogenic response is relevant to human. Category 3: "*possible human carcinogen*" - limited evidence of carcinogenicity in humans and animals with some supporting data (genotoxicity, DNA reactivity, metabolism, mechanism of action, structure/activity relationships, etc.); or inadequate evidence of carcinogenicity in humans, sufficient evidence in animals, inadequate evidence on relevancy; or strong evidence for lack of carcinogenicity in humans (several adequate studies showing no association between exposure and cancer), sufficient evidence in animals with positive evidence that animal studies are relevant to humans. Category 4: "*equivocal evidence for carcinogenic activity*" - inadequate evidence of carcinogenicity in humans and limited evidence of carcinogenicity in animals with little or no evidence of relevancy to human. Category 5: "*evidence inadequate for classification*" - inadequate human evidence and inadequate or suggestive evidence of noncarcinogenicity in animals. Category 6: "*carcinogenic activity in animals; probably not a human cancer hazard*" - human evidence inadequate or suggestive of noncarcinogenicity; sufficient or limited animal evidence with evidence from humans or experimental animals indicating that animal studies are not likely to be predictive for humans. Category 7: "*carcinogenic activity in animals; considered not a human cancer hazard*" - human evidence inadequate or suggestive of noncarcinogenicity; sufficient or

limited animal evidence shown not to be predictive of a human response by strong evidence from human or experimental animal studies. Category 8: "*evidence of noncarcinogenicity*" - valid information exist indicating that the agent lacks carcinogenic activity.

EPA's weight-of-evidence categories in the 1996 proposed guidelines are very different from those describe in the 1986 guidelines; the categories are also simpler and more straight forward than those described in the working paper (U.S. EPA, 1992a). All the evidence is brought together in a hazard narrative that would include the hazard descriptor or category (*known/likely*, *cannot be determined*, and *not likely*) designated by route of exposure. The *known/likely* category includes agents for which there is sufficient causal evidence of carcinogenicity from epidemiologic studies as well as agents for which there are no human data. The guidelines include a subcategory, *treat as if known*, for evidence showing "plausible", but not definitive, causal association between exposure to an agent and human carcinogenicity and strong evidence from animal studies. The question is whether the descriptor, *known*, should be used for agents for which the evidence from humans studies is not definitive. Some exceptions should be considered, such as chromium IV, which should include all chromium IV compounds, arsenic and its compounds, and other similar agents. The guidelines place the descriptors, *known* and *likely,* in the same category. Considering the degree of uncertainty in the levels of evidence for the two descriptors, should these two descriptors be placed in the same category or put into two separate categories? The descriptor, *not determined*, encompasses evidence levels ranging from "suggestive" or "equivocal" to "no data." There is no clear category or descriptor for agents in which the evidence is not sufficient for a *likely* descriptor, but is believed to pose a hazard because the evidence shows a strong reason for concern for potential carcinogenicity (e.g., only one available study in one sex and species showing carcinogenicity and genetic toxicity data are limited). Should the guidelines include an additional category for suggestive evidence or for agents showing strong reasons for concern?

5.1.5. Maximum Tolerated Dose

The concept of maximum tolerated dose (MTD) is incorporated into risk assessments to ensure that animal studies evaluated for carcinogenicity attain the sensitivity to detect a carcinogenic response given that only a limited number of animal can be used for testing. Testing at doses considered to be too high or too low serves as a reason for rejecting studies for risk assessment purposes. Therefore, a discussion on the role of maximum tolerated dose (MTD) in risk assessment deals to a large degree with the science policy of carcinogen testing. The primary basis for testing at the MTD is to minimize the chances a "carcinogen" remaining undetected and to compensate for using a small number of animals because of limited statistical power to detect

a significant increase in tumor incidence (Carr and Kolbye, 1991). However, the highest doses tested should cause no undue toxicity that would confound the interpretation of the study.

Haseman (1985) noted that some of the confusion about MTDs centers around the definition of an MTD. Haseman and Lockhart (1994) presented the definition used as the basis for the dose selection (selecting the top dose) process in the National Toxicology Program's (NTP) testing protocol. The NTP's definition came from Sontag et al. (1976) and is defined as the "highest dose of the test agent during the chronic study that can be predicted not to alter the animals' normal longevity from effects other than carcinogenicity." Haseman (1985) stated that the Sontag et al (1976) definition included the caveat that, in the subchronic study, the doses selected should not cause more than a 10% decrement in body weight compared with appropriate controls. Carr and Kolbye (1991) included a weight loss of no more than 10% in their definition of MTD. Currently, in NTP studies, body weights do not factor as prominently in selecting the MTD as do other signs of toxicity, such as development of nonneoplastic lesions and their prediction to be life threatening during long-term treatment (Haseman, 1985). Haseman and Lockhart (1994) stated that the NTP prefers to use the term "minimally toxic dose" for its dose selection process. Overall, the selection of the top dose for long-term NTP studies includes data on body weight, survival, histopathology, clinical and pharmacologic signs, and metabolism and disposition information obtained from 90-day studies (Haseman and Lockhart, 1994). Some toxicity at the top dose in a long-term study confirms that the animals have been sufficiently challenged (Haseman, 1985).

The definition of MTD as presented in EPA's working paper (U.S. EPA, 1992a) was "a dose which is estimated to produce some minimal toxic effects in a long-term study (e.g., a small reduction in body weight), but should not shorten an animal's life span or unduly compromise normal well-being except for chemically induced carcinogenicity."

In attempting to maximize the ability to detect weak carcinogens, other problems of organ toxicity and possibly tumor promotion can distort the interpretation of the results (Carr and Kolbye, 1991). Therefore, one criticism of using the MTD to select the top dose for carcinogenicity studies is organ toxicity, which is manifested by cell killing and regenerative hyperplasia, producing conditions having no relevance to humans exposed to lower doses (Haseman and Lockhart, 1994).

"Metabolic overloading" and/or "secondary carcinogenesis" may be reasons for rejecting the MTD concept of dose selection (Haseman, 1985). Carcinogenesis associated with metabolic overloading is caused by saturation of the detoxification mechanism. Secondary carcinogenesis is due to induction of excessive nonspecific tissue damage. Haseman (1985) stated that a

distinction should be made between saturation of the activation pathway and saturation of the detoxification pathway. He also stated that, if increases in tumor incidences at high doses are to be explained away because of metabolic overload or secondary carcinogenesis, a direct cause and effect relationship should be established between these factors and tumor induction (i.e., one must show how the overload produces carcinogenic effects). It is not enough just to show that the overload occurs.

Another criticism of testing at the MTD is that positive evidence for carcinogenicity would not have been obtained for two-thirds of the NTP carcinogens if the top dose had been excluded (NTP, 1992). According to Griesemer (1992), this conclusion was based on a misinterpreted report by Haseman (1985). Haseman (1985) concludes that, if the high dose had been reduced from the MTD to ½MTD, then two-thirds of the carcinogenic effects in feeding studies would be eliminated. Eight of 13 chemicals were judged to be carcinogenic based only on effects at the top dose. On the other hand, some equivocal results would have been regarded as real carcinogenic effects if the top dose had been excluded. After evaluating the results of 216 chemicals tested by the National Cancer Institute (NCI) and NTP, Haseman and Lockhart (1994) concluded that, even without the top dose, two-thirds of the carcinogens would have been detected, but not some of the site-specific effects.

A third concern regarding the MTD concept is the overestimation of the high dose resulting in excessive toxicity and mortality. Haseman (1985) countered this argument by noting that the standard dosing protocol for NTP studies is to include three doses (MTD, ½MTD, and ¼MTD), insuring that if unanticipated chronic toxicity is seen at the top dose, the two lower doses provide a margin of experimental safety. In addition, in 31 NTP feeding studies, survival of high-dose animals generally exceeded that of controls, and body weights were reduced in some experiments, but not enough to suggest that the MTD had been exceeded (Haseman, 1985).

Carr and Kolbye (1991) recommended that the MTD be replaced by a minimally toxic or the highest subtoxic dose that can be tolerated over a long period of time. The high dose would not produce life-shortening, weight loss, or demonstrable organ or tissue toxicity. However, long-term studies often serve a dual purpose – documentation of carcinogenicity as well as chronic toxicity. If no effects are produced at the highest dose in long-term toxicity studies, then chronic toxicity cannot be documented. Carr and Kolbye (1991) further stated that nonneoplastic toxicity could be assessed by testing at high doses achieving minimal toxicity; they further stated that carcinogenicity may be grossly exaggerated at minimally toxic doses. OSTP (1985) stated that pharmacokinetics and metabolism should be included in the dose selection process. Butterworth et al. (1991)

proposed that cell proliferation in 90-day studies be used as an additional criterion for selecting doses in long-term studies.

OSTP (1985) stated that the high dose should maximally enhance the sensitivity of the test without introducing qualitative distortion of the results. In its 1986 guidelines, EPA (U.S. EPA, 1986) stated that "long-term animal studies at or near the maximum tolerated dose (MTD) are used to ensure an adequate power for the detection of carcinogenic activity." The agency further noted that carcinogenic responses at doses exceeding the MTD should be review carefully as to their relevancy to humans. In the 1996 proposed guidelines, EPA (U.S. EPA, 1996b) asserted that failure to reach adequately high doses reduces the sensitivity of the study to detect a response, whereas overt toxicity due to excessive doses raises questions about the specificity of the response, whether it is related to exposure or to secondary toxic effects. Findings that confound the interpretation of studies are significant mortality not due to cancer; body weight decrements greater than 10%; and significant toxicity manifested by clinical signs, clinical chemistry and hematological changes, organ weight changes, and gross and histopathologic changes. The 1996 proposed guidelines presented general guidance for evaluating the dosing protocol in animal studies. The guidelines also asserted that studies showing excessive toxicity "are generally not suitable for risk extrapolation." It appears that the guidelines have adequately addressed this issue and provided guidance for evaluating dosing issues in carcinogen risk assessment. It should be noted that long-term animals studies serve a dual purpose, that of documenting systemic toxicity as well as carcinogenicity. Because of the extensive resources required to conduct these studies, it may be necessary to compromise on certain issues. Therefore, definite toxicity at the high dose is necessary to document the two types of responses in a study.

5.1.6. Dosimetry: Pharmacokinetic or Toxicokinetic Modeling

Pharmacokinetic models can be used in both cancer and noncancer assessments to estimate internal and delivered doses and to extrapolate doses across species when empirical data are available. EPA's 1986 cancer guidelines state that "In the absence of comparative toxicological, physiological, metabolic, and pharmacokinetic data for a given suspect carcinogen, extrapolation on the basis of surface area is considered to be appropriate because certain pharmacological effects commonly scale according to surface area." The 1996 proposed cancer guidelines takes the position that available data are evaluated to reach a measure of internal or delivered dose. The inhalation RfC methodology recommends pharmacokinetic models for route-to-route extrapolations and estimating human equivalent doses (U.S. EPA, 1994a). These models are an improvement over using the default dose scaling method for cancinogen dose-response assessments or the default RDDR or RGDR methodology for inhalation toxicants. EPA has not

established guidance for determining when data are sufficient to apply pharmacokinetic models to estimate internal or delivered doses.

The data base on pharmacokinetic models is too extensive to be discussed here. These models are applied in risk assessment on a case-by-case basis.

5.1.7. Dosimetry: Default Dose Scaling Methods

Dose-response data from human studies are seldom available for quantitating risk due to exposure to environmental substances; therefore, it is necessary to use the data available from animal studies. The human dose capable of producing a response similar to that of experimental animals can be estimated by one of two ways: case-specific pharmacokinetics data or in the absence of data, a default method, which extrapolates or scales animal doses to equivalent human doses. In the 1986 carcinogen risk assessment guidelines EPA's default dose-scaling method was based on evidence that "certain pharmacological effects commonly scale according to surface area" (U.S. EPA, 1986) and the proportionality of body surface area to the ⅔ power of the body weight (U.S. EPA, 1980, 1986). The resulting equation for calculating the human equivalent dose is as follows: HED = animal dose (mg/day) × (human weight (kg)/animal weight (kg))^{⅔} or HED = animal dose (mg/kg/day × (animal weight (kg)/human weight (kg))^{⅓}. Scaling based on body surface area is considered to be the most accurate method of scaling (Calabrese, 1991); however, measuring individual surface areas is inherently difficult (Calabrese, 1991) and, therefore, not feasible. Calabrese et al. (1992) stated that dose adjustments using surface area results in similar blood levels of the substance reaching potential target organs, but some principal causes for intraspecies difference in response are not addressed (metabolic and pharmacodynamics factors). Travis and White (1988) analyzed data on direct acting carcinogens and reported that interspecies dose scaling should be based on body weight raised to the ¾ power.

In the proposed guidelines (U.S. EPA, 1996b), EPA adopted the default dose scaling procedure based on body weight raised to the ¾ power. The basis for the change in the default scaling method can be found in EPA's 1992 Federal Register report (U.S. EPA, 1992b). This scaling method takes into account both toxicokinetic (area under the concentration curve (AUC) and toxicodynamic (mechanistic data) parameters such that lifetime equivalent doses required to produce lifetime equivalent responses among species are predicted. Calabrese (1991) also presented a detail discussion on cross-species extrapolation. EPA considers dose scaling based on body weights to the ¾-power is a more scientifically defensible default method than the ⅔-power scaling, and it is amenable to incorporation of case-specific data (U.S. EPA, 1996b).

5.1.8. Low-Dose Extrapolation (Estimation of Risk at Low Doses)

Gad and Weil (1986) described three basic steps to low-dose extrapolation: (1) identifying the dose-response data points, (2) selecting a mathematical model to extend to observations from the experimental region to the region of concern, and (3) making a basic assumption about the nature of the dose-response relationship at extremely low doses. Steps 2 and 3 should be reversed; assumptions about the nature of the dose-response relationship at low doses should precede selecting the model to extend the data to low doses. Gad and Weil (1986) stated that the assumption about the low-dose region concerns the question of threshold. EPA's default positions are related to the nature of the dose-response relationship(linear, nonlinear, or both (U.S. EPA, 1996b). A threshold concept is encompassed within the nonlinear dose-response relationship.

Gad and Weil (1986) presented some arguments for and against the existence of a threshold for carcinogens. The arguments against a threshold are as follows: (1) a single molecule of a chemical can mutate a cell, *in vitro*; (2) the presence of other agents in the environment may act as a promoter for the agent of concern or saturate the existing defense mechanism; (3) a threshold would preclude a linear dose response; and (4) there may be thresholds for some or most individuals in a population, but, not for all individuals, i.e., there are no exposures that are absolutely safe for absolutely everyone. The arguments for a threshold are as follows: (1) most carcinogens and mutagens exhibit a dose-response relationship, which show an apparent threshold for some agents; (2) toxicity, including carcinogenicity, is a result of pharmacokinetic processes (absorption, distribution to tissues, reaction with cellular components, adaptation and repair by molecular and cellular components, and clearance from the body by metabolism and/or excretion), which are linear only within particular ranges, and metabolic thresholds may lead to disproportionate increases in toxicity at certain dose levels; (3) a biological threshold is suggested based on probabilistic grounds, because the probability of a "hit" by a carcinogen producing an initiating or promoting effect is low; (4) as the dose decreases the time-to-tumor increases, which could eventually result in a time-to-tumor exceeding the life span of the exposed population; and (5) there are physicochemical factors (or mechanisms) that cause some agents to be carcinogenic above certain doses only. Cohen (1981) stated that the justification for the linear-no-threshold model leaves much to be desired and the evidence does not support it at low doses. Wilkinson (1987) stated that extrapolation below the experimental range often encompasses four to five orders of magnitude and that most areas of science would not attempt such an extrapolation. He further stated that, from a practical standpoint, thresholds must exist.

Low-dose extrapolation is an essential element of quantitative risk assessment, because exposure levels encountered in epidemiologic (usually occupational exposure) and animal studies

do not achieve the low level environmental exposures experienced by the general population. To achieve statistical significance of response to such low levels, animal studies would require vast numbers of animal and economical resources so as to make such studies unfeasible. Consequently, mathematical models have been developed to extrapolate experimental doses to low level environmental exposures. The conceptual basis for the low-dose extrapolation is intermixed with the linear nonthreshold concept of carcinogenesis. In one of it principles, OSTP (1985) stated that "mechanistic considerations such as DNA repair and other biological responses, in general, do not prove the existence of, the lack of existence of, or the location of a threshold for carcinogenesis." OSTP (1985) further noted that no single mathematical model can be used for low-dose extrapolation, but the model chosen should be consistent with the evidence. However, when data are limited, models incorporating low-dose linearity are preferred. In its 1986 guidelines, EPA's default procedure for low-dose extrapolation was the linearized multistage model (a mathematical curve-fitting model based on a nonthreshold concept). The Agency noted in its guidelines, that the model selection should be consistent with the evidence. The linearized multistage model estimates the upper limit on the risk at low doses; however, according to the EPA (U.S. EPA, 1986), the true risk is unknown and could be as low a zero.

In the 1996 proposed guidelines, dose-response assessment is conducted in two steps: modeling within the experimental range to determine the LED_{10} or another point of departure and extrapolation below the point. Biologically-based or case-specific models are used when sufficient data are available for both steps, otherwise curve-fitting models are used for the experimental range and a straight line extrapolation from the point of departure to the origin is used for the linear approach; the margin of exposure procedure is used for the nonlinear approach. The linear approach produces a probabilistic risk of cancer and the nonlinear approach produces a ratio between the dose at the point of departure and actual environmental exposures.

The technical review workshop reviewers (Eastern Research Group, 1994) noted that the linearized multistage or other low-dose extrapolation procedures are generally inappropriate for "extrapolating risk from the upper-bound confidence intervals and dose from the lower-bound confidence intervals....."; the reviewers agreed with EPA's proposal to perform a simple straight line extrapolation to the zero response, use the linear procedure when sufficient data for applying other procedures are not available (default position), and use nonlinear procedure only when extensive data are available to support a nonlinear procedure.

Although the risk manager determines whether the margin of exposure is adequate, the risk assessor should recommend an adequate margin of exposure and the basis for the

recommendation. The 1996 proposed guidelines discussed some issues that should be considered in determining what is an adequate margins of exposure. There are other issues, in addition to those discussed in the guidelines, that should be considered.

The 1996 guidelines discusses the use of factors (as used for RfD and RfC derivation) applied to the point of departure in the analysis of the margin of exposure (no less than 10 for intraspecies variability and 10 for interspecies sensitivity). It should be pointed out that human equivalent doses estimated by either default or toxicokinetic models are incorporated into general curve-fitting or chemical-specific procedures to determine the LED_{10}. Therefore, elements for species sensitivity are already incorporated into the LED_{10}. To derive an RfD, experimental animal doses are usually adjusted by a factor of no more than 10 (a factor of 3 is applied for RfC derivation); therefore, a factor of "no less than 10-fold", as recommended in the proposed cancer guidelines (U.S. EPA, 1996b), may be too large to account for interspecies sensitivity. Additionally, applying a factor of "no less than 10-fold", as recommended by the cancer guidelines (U.S. EPA, 1996b), to account for human variability, implies that the LED_{10} (95% confidence limit on dose) is exclusive of the sensitive population. The factors applied to margin of exposure analysis are similar to the uncertainty factors applied to RfD/RfC derivations. The guidelines did not explain how an analysis of intraspecies and interspecies considerations would be applied to the margin of exposure analysis. Nevertheless, a margin of exposure of 100 would not be protective of sensitive populations, because the environmental dose would correspond to a 10% response in the sensitive population whether the response measured is tumor induction of a precursor lesion. Applying factors of 10 each to account for interspecies and intraspecies variability does not correspond to reduction in the response level, but only for intraspecies and interspecies differences.

5.1.9. Dose-Response Model Selection

In the 1996 guidelines, EPA proposes that curve-fitting models be used to model data in the experimental range when biologically-based or case-specific data are not available. EPA, however, did not recommend a default curve-fitting model to be used. Perhaps a default curve-fitting model should be used except when data is best described by another model. Several mathematical models have been described by Gad and Weil (1986) and Johannsen (1990). These models have been applied to low-dose extrapolation, but could be used to model data in the experimental range. Gad and Weil (1986) discussed seven models and Johannsen (1990) discussed ten. The various models can be placed in the following categories (examples of each model): linear, mechanistic (one-hit, multihit, multistage, linearized multistage), tolerance distribution (probit, logit, and Weibull), time-to-tumor (Weibull distribution), and biologically

motivated (Moolgavkar, Venzon, and Knudson(MVK)). Johannsen (1990) also ranked the conservativeness of some of the mathematical models applied to low-dose extrapolation as follows: one-hit > linear > multistage » Weibull > multihit » logit > probit.

5.1.10. Uncertainty Analysis

NRC/NAS (1983) noted that one of the inherent limitations of risk assessment is the pervasive uncertainty due to incomplete data sets and the estimates of types, probability and magnitude of health effects associated with exposure to chemical agents. The NRC/NAS (1994) Committee placed considerable emphasis on uncertainty analysis in risk assessment. The Committee's main concern regarding EPA's lack of quantitative uncertainty analysis in its risk assessment process appears to be the difficulty in determining the degree of conservatism in the risk estimate. The Committee recommended that EPA conduct a formal uncertainty analysis, identify errors of either overestimation or underestimation, and develop guidelines for quantifying and communicating uncertainty.

In the 1996 proposed guidelines EPA discussed two types of uncertainty: model and parameter. Model uncertainty, which deals with biological questions, is described qualitatively or by presenting alternative results when more than one extrapolation model can describe a particular data set. Parameter uncertainty, which deals with statistical or analytical measures of variance in data set or estimates, are described quantitatively.

5.2. Noncancer Risk Assessment

5.2.1. General Guidelines for RfD/RfC Derivation

The NOAEL/LOAEL approach to deriving risk values for noncancer effects have traditionally made little use of the biological data available for chemical assessments. Very little quantitative use is made of the wealth of information available on some agents, because the assessment, although based on scientific judgment, is reduced to selecting a critical effect (one effect seen at the lowest dose) and applying predetermined uncertainty factors to the largest dose at which the critical effect does not occur. Biological data on mode of action data and dose-response relationships have traditionally been exempt from assessments of noncancer effects. Few epidemiologic studies are available for assessing noncancer studies in humans, and the more detailed scientific analysis of animal models to determine the likely occurrence of a response in humans have not received the attention as in cancer assessments. Part of the problem is that noncancer data are often presented in a format that cannot be readily incorporated into dose-

response models, but could be readily used for the NOAEL/LOAEL approach for noncancer assessments

Reference values are derived by applying uncertainty factors to NOAELs and LOAELs to account for the inherent scientific uncertainty (U.S. EPA, 1991b). They address issues of variability in the human population, the existence of sensitive subpopulations, and the differences (pharmacokinetic and pharmacodynamic) between humans and animals (Calabrese et al., 1992). An uncertainty factor of 10 for interspecies variations (U.S. EPA, 1993) is intended to account for toxicokinetic and toxicodynamic uncertainties (U.S. EPA 1991b). Dourson and Stara (1983) analyzed the 490 probit, log-dose slopes for acute lethality presented by Weil (1972) and concluded that a 10-fold uncertainty factor to account for intraspecies variability was reasonable when chemical specific data are not available. Dourson and Stara (1983) also concluded that a 10-fold decrease in animal dose was adequate to adjust to human doses in the absence of chemical-specific data if the assumed dose equivalence based on surface area is correct. They noted a need for additional investigation on this subject.

Calabrese et al. (1992) compared dose adjustments based on surface area with the application of uncertainty factors. They found that for the mouse, a tenfold uncertainty factor would not be as protective as the surface area adjustment of the dose; whereas a tenfold uncertainty factor for the rat would be more protective than surface area adjustment of the dose. They concluded that humans are likely to be more protected if the critical study for deriving a reference value is based on a large-sized animal. According to Calabrese et al. (1992), EPA's argument that the animal-to-human uncertainty factor incorporates the surface area normalization (1) has inherent toxicologically based contradictions, (2) has an inadequate theoretical foundation, (3) differentially protects the public according to the animal model, and (4) includes no interspecies uncertainty factor for non-surface area normalization for the mouse.

Gaylor (1983) noted that the use of safety factors (uncertainty factors) results in an uneven control of risk depending on the number of animals used in a study. As an example, he noted that if the same percentage of positive responses are seen in a study using 20 animals per dose and one using 60 animals, the study using only 20 animals (lower statistical power resulting in a higher NOAEL) would receive the same uncertainty factor and result in a higher reference value. This example is critical of using uncertainty factors applied to experimental no effect levels. Gaylor (1983) recommended that uncertainty factors should be applied to a dose that would results in a risk below that of the no effect level (e.g., ED_{01}, the dose at which there is a 1% response compared with controls). Because this dose is not easily determined experimentally, Gaylor (1983) suggested that a curve-fitting model be used to determine the upper confidence limit on the risk

at ED_{01}, which could be used as an estimate of the uncertainty reflected by the sample size. EPA uses the lower confidence limit on dose instead of the central tendency in the benchmark approach.

Another major criticism of the present method for deriving reference values is that the NOAEL/LOAEL approach does not use all of the dose-response information. In addition, the NOAEL must come from the experimental data, and dose spacing can greatly influence the outcome of the evaluation (U.S. EPA 1991b). The benchmark procedure for deriving reference values has become a primary focus of the Agency in addressing some of the inherent limitations of the NOAEL/LOAEL approach (U.S. EPA, 1991b). The benchmark approach involves using a dose-response (curve-fitting) model to estimate the lower confidence limit on the dose (generally the LED_{10}, LED_{05}, or LED_{01}). This procedure is similar to that described by Gaylor (1983) and the issues are discussed in more detail in section 5.2.2.

5.2.2. Benchmark vs NOAEL/LOAEL Approach to Dose-Response Assessment

The BMD approach to noncancer risk assessment has the potential of incorporating more biological data into the assessments and is also be more comparable to cancer assessments. The robust databases available for some chemical agents will result in reductions in uncertainties and increases in the confidence in the risk values obtained.

One of the major issues regarding benchmark dose and the NOAEL/LOAEL approaches is the application of uncertainty factors. When animal data are used to derive reference values, two primary uncertainty factors are applied, one accounting for intraspecies variability and the other accounting for interspecies sensitivity. Comparison of the LED_{10} with the ED_{10} (central tendency) may suggest that it is possible to reduce or eliminate the tenfold uncertainty factor accounting for intraspecies sensitivity. The 95% confidence limit on dose should account for some of the population variability of a response, but the dose for supersensitive individuals may fall outside the range of the 95% confidence limit. When a large segment of the population fall with the sensitive subpopulation as may be the case for developmental (fetus or young children) or reproductive effects (women or men during reproductive age), the BMD analysis can be based on these specific endpoints. When studies with adults show strong sex specificity, the BMD should be based on the most sensitive sex. Under these conditions, the RfD/RfC can be directly derived for the sensitive population, thus eliminating the need for an intraspecies uncertainty factor.

The uncertainty factor accounting for interspecies sensitivity may be reduced by using dose scaling to derive human equivalent doses similar to the method used for dose-response

assessments for cancer endpoints. The RfC methodology uses dosimetric adjustments for inhalation of particles and gases; this adjustment accounting in part for differences in interspecies sensitivity, resulted in a reduction of the uncertainty factor from 10 to 3 (U.S. EPA, 1994a; Jarabek, 1995a). EPA has a default dose scaling method ($BW^{3/4}$) when chemical-specific data on pharmacokinetics and metabolism are not available for carcinogen assessments. By using a similar method for noncancer assessments, the uncertainty factor accounting for interspecies sensitivity could be reduced. The precedence for using this method is based on the similarity of the benchmark approach to the nonlinear approach for cancer assessments.

Uncertainty factors accounting for intraspecies and interspecies sensitivities should not assume a reduction in the response level for the resulting doses. Therefore, an additional uncertainty factor should be incorporated into derivation of RfDs/RfCs, that of extrapolating from the LED_{10} to a safe dose, or a dose expected not to have adverse effects. The LEDs, which approximates the 95% statistical bound on the lowest-effect level, are effect levels whether they are based on 1, 5, or a 10% response, and should not be considered as no-effect levels. Therefore, a factor accounting for extrapolation to a non-adverse effect should be applied to the LED_{10}. Just as the slope of the dose-response relationship at the point of departure should be considered in deciding on an adequate margin of exposure (U.S. EPA, 1996b), the slope at the BMD should also be considered in the extrapolation to a dose expected to have no adverse effects.

The BMD has been compared with the NOAEL (Beck et al., 1993; Allen et al., 1994), and it is questionable whether this is a valid comparison. The benchmark method can be compared with the NOAEL/LOAEL method for deriving reference values, but the value of the BMD derived from a study should not be compared with the value of the NOAEL identified in the study. The comparisons may be invalid, because the BMD is a statistically derived value, whereas the NOAEL is empirically derived and is dependent on the statistical power of the study. However, it is valid to compare RfDs/RfCs derived by the NOAEL/LOAEL approach with the value derived using the benchmark approach.

Barnes et al. (1995) addressed a number of issues pertaining to using the benchmark approach to calculate BMDs and reference values. Some of the issues discussed concerned selection of dose-response models, use of a default dose-response model, analysis of non-quantal data, characterization of uncertainties in benchmark estimation, and application of uncertainty factors to the BMD.

The benchmark approach is a different perspective and could be an improvement over the NOAEL/LOAEL approach to deriving risk values for noncancer endpoints. There are advantages and disadvantages to using the benchmark approach; the disadvantages are due in part to the testing protocol, which confirm with the current NOAEL/LOAEL method of deriving noncancer risk values. The implementation of the benchmark approach may result in changes in the testing protocols that would provide data more easily incorporated into the benchmark methodology. The major advantage of the benchmark approach is that more of the available data can be used to derive risk values.

5.2.3. Developmental toxicity: maternal/developmental toxicity

Developmental toxicity effects at dose levels that result in pronounced maternal toxicity may be difficult to interpret. However, for risk assessment whether a developmental effect is secondary to maternal toxicity or not, does not affect the selection of the NOAEL. One approach for ranking substances according to their relative maternal and developmental toxicity involves the calculation of the ratio of the adult toxic dose to the developmental toxic dose (A/D ratio) (Johnson and Gable, 1983). However, comparison of A/D ratios for 14 chemicals found little agreement between four species (Daston et al., 1991).

5.2.4. Developmental toxicity: functional toxicity

Functional developmental toxicity endpoints may be used for establishing the NOAEL when these endpoints are found to be the adverse effect occurring at the lowest dose. Support of the use of behavioral assessment has been shown for human lead exposure data (Annau, 1990). However, debate still exists on weighting tests within the battery of functional neurotoxicity tests and whether observed effects are due to neurotoxicity alone or as part of overall developmental toxicity (Tyl and Sette, 1990).

5.2.5. Neurotoxicity: Endpoint Determination and Dose-response Assessment

Similar to most toxicants, hazard identification for neurotoxicants is complicated by uncertainty regarding the definition of adverse effect. Starting with the NAS definition of adverse effect, various modifications and refinements have been suggested including that of the EPA test guidelines (U.S. EPA, 1985) which states that a neurotoxic effect is "an adverse change in the structure or function of the nervous system following exposure to a chemical agent" and that of Spencer and Schaumberg (1985) defining neurotoxicity as "a consistent pattern of neurological dysfunction in humans, comparable dysfunction in animals, and reproducible lesions in animals/humans that are related to the neurobehavioral dysfunction expressed". Tilson (1990) suggested the following criteria for defining neurotoxicity: (1) side effects or overdose (unwanted

effects), (2) decreased ability to function fully or [provide] compensation in order to function normally, and (3) an alteration that diminishes the ability to survive, reproduce, or adapt to the environment.

Selecting the most appropriate toxicity endpoint for dose-response or exposure-effect determination remains a point of concern. Endpoint selection may be affected by various elements such as species, gender, and dosing regimen. Additionally the selection of an endpoint appropriate for the risk assessment process is often subjective and dependent upon the definition of "adverse effect". Neurotoxic effects reported in humans have been categorized by Tilson and Cabe (1978) and Reiter (1987) as: (1) sensory disorders, (2) cognitive disorders, (3) changes in CNS excitability, (4) autonomic dysfunction, (5) motor disorders, (6) sleep disturbances, (7) affective disorders, and (8) physiological alterations. Stanton and Spear (1990) identified more general groups for assessment of neurotoxicity: sensory, motivational, cognitive, motor, and social functional. Many of these categories may be functionally interrelated and some level of integration would be required for expression of the neurotoxic effect.

The U.S. EPA (1994b) notes that the selection of a toxicity endpoint is dependent upon the level of nervous system organization being investigated - biochemical, anatomical, physiological, or behavioral. General categories of neurotoxicity endpoints include behavioral (learning and memory, altered behaviors, etc.), neurochemical (changes in synthesis, release and uptake of neurotransmitters, alterations in membrane-bound enzymes relevant to neuronal activity, etc.), neurophysiological (changes in nerve conduction parameters, etc.), and structural (accumulation, breakdown, or rearrangement of structural elements, etc.).

Because of the subtleties of neurotoxic effects, and the uniqueness of the nervous system and its responses to toxic insults, the use of biomarkers and dose has been suggested as a method for hazard identification (Gaylor and Slikker, 1990). Specifically, this method involves four steps, the first three of which would pertain to hazard identification: (1) establishing a mathematical relationship between effect or biomarker and dose, (2) determining the distribution of the individual measurements of effect or biomarker around the dose-response curve, (3) establish a level of the effect or biomarker that would be considered abnormal or adverse, and (4) assess the proportion of individuals exceeding the adverse or abnormal level of the effect or biomarker as a function of dose.

Currently, dose-response assessments are made with the assumption that a threshold exists for a toxicologic response and that below this threshold there will be no significant response.

As previously noted, however, the commonly used NOAEL/LOAEL approach may not provide a true measure of the no-adverse-effect-level; there may be toxic responses below the adopted NOAEL or the NOAEL may actually be an overestimation of the true response level. Furthermore, this method does not make use of the complete dose-response curve. Alternative methods have been described (Gaylor and Slikker, 1990; Dews, 1986; Glowa and Dews, 1987; Glowa et al., 1983; Crump, 1984b) that make use of complete dose-response data and mathematical functions to estimate the variability in exposures or dose and effects. Because dose-response curves may not be linear for all chemicals, these alternative methods may provide more accurate and definitive assessments of the dose-response relationship.

5.2.6. Interspecies Extrapolation

Although not unique to neurotoxicants, the issue of interspecies extrapolation presents special problems regarding toxicity endpoints of cognitive function and behavior. For example, there is no direct animal counterpart for psychometric IQ but conceptual analogs can be developed to assess changes in neurobehavioral and cognitive functions in animals (Winneke, 1992). When extrapolating from animal data, quantitative assessments must address three basic problems: 1) identification of a relevant adverse effect, 2) high-to-low dose extrapolations, and 3) assessing dose equivalency (Rees and Glowa, 1994).

5.2.7. Reversibility

Although hazard identification principles are similar for neurotoxic effects as they are for other adverse effects, the phenomenon of reversibility requires special consideration. Unlike many organ systems, the nervous system exhibits considerable redundancy and plasticity but also lacks the repair potential of most tissues. The redundancy and plasticity often results in apparent recovery (reversibility) from a toxic insult when, in fact, damage has occurred that may become manifest at a later time or under different conditions.

5.2.8. Delayed Neurotoxicity

Organophosphate induced delayed neurotoxicity (OPIDN) guidelines have been established for pesticide assessment (U.S. EPA, 1991d). OPIDN assessment focuses on biochemical (e.g., inhibition of acetylcholinesterase and neurotoxic esterase) and behavioral changes that are known to occur several days to several weeks after acute and short-term exposure to organophosphates.

6. KEY ISSUES

A large number of issues pertinent to cancer and noncancer risk assessment have been discussed in this report. The following is a list of the most pertinent or key issues.

(1) Target organ concordance – Rodent species are most often used as surrogate for human carcinogenicity studies. Although there are species specificity in the targets affected by carcinogenic agents, multiple targets in one species often indicate that an agent will elicit a response in another species, but not necessarily at the same sites. Carcinogenic agents eliciting responses in multiple animal species at multiple target are more likely to be associated with carcinogenicity in human studies, no matter the sites affected. Rodent species have anatomical sites that have no analogues in humans (e.g., forestomach and Zymbal's gland) that are the targets for carcinogens. Sometimes these sites may be the only target in a rodent study. Is carcinogenicity occurring only in organs having no analogues in humans predictive of human cancer risk?

(2) Genotoxic carcinogens – Mode of action information is an important aspect of hazard assessment and is used to determine the dose-response procedure to apply to a particular agent, and genetic toxicity is pivotal is determining mode of action. There are many tests available for determining the genotoxicity of an agent. Which tests or endpoints are considered the most important for determining genotoxicity (e.g., DNA adduct formation, positive response in the *Salmonella* test, micronucleus test, dominant lethality, etc.) and how much weight should be given to a particular genotoxicity test. What are the criteria for determining genotoxicity?

(3) Role of cell proliferation in carcinogenesis – Induction of cell proliferation has been implicated in the mode of action of nongenotoxic carcinogens. It has been postulated that cell proliferation increases the probability of a mutational event. However, mutations have not been detected as a result of cell proliferation. Therefore, what is the role of cell proliferation in carcinogenesis? Could the reason for not detecting carcinogenesis at doses below those causing cell proliferation be due to the limited detection power of the rodent bioassay for weak carcinogens. Is cell proliferation alone sufficient for cancer induction?

(4) Route-to-route extrapolation – According to the 1996 proposed cancer guidelines, the only requirement for route-to-route extrapolation of hazard is to show absorption by another route to give an internal dose There are other factors that may determine whether cancer would be induced by another route. Because of first pass effects, metabolism and tissue distribution may

be different. The guidelines also stated that the hazard classification is route specific. Why allow route specific hazard classification along with liberal requirements for route extrapolation? A small change in the hazard categories could address the issue of route-to-route extrapolation of hazard. There is a need for a "*suggestive*" or "*evidence shows reason for concern*" category. Evidence of absorption of an agent by another route to give an internal dose is definite reason for concern for the second route. There are cases in which route-to-route extrapolation of cancer hazard and potency can be conducted with less uncertainty. EPA (U.S. EPA, 1996a) concluded, in its dose-response assessment of PCBs, that a cancer risk by skin contact and inhalation exposure is possible. This conclusion was based on absorption by both route, but most importantly was the slow metabolism rate and accumulation of PCBs in fat, would cause a slow release of PCBs over a prolonged period of time. Therefore, the slow metabolism and fat accumulation of PCBs would reduce the uncertainty associated with first pass effects.

(5) Curve fitting in the experimental range – The 1996 proposed cancer guidelines states that curve-fitting models can be use as the default procedure for modeling data within the range of experimental observation. Curve-fitting models can be used also for deriving BMDs for noncancer assessments. A variety of curve-fitting models available, including the multistage model, most of which can be used to model the data and estimate the LED_{10}. Unless there are compelling reasons to recommend another model, a default curve fitting model could be applied to experimental data, which would provide consistency in the risk assessments.

(6) Low-dose extrapolation: margin of exposure procedure for cancer assessments – A straight line approach is used for extrapolation below the point of departure when data are not available for applying a case-specific model and the mode of action suggest a linear procedure. When the mode of action shows evidence for a nonlinear procedure the margin of exposure approach is used for low-dose extrapolation. The risk manager determines whether the margin of exposure is adequate based on the recommendations of the risk assessor. The recommendations are dependent on a number of factors including shape and slope of the dose-response curve at the point of departure. There are, however, a number of questions concerning the margin of exposure analysis. What is an adequate margin of exposure protective of sensitive individuals? Would an LED_{10} based on a 97 or 99% rather that a 95% confidence limit encompass sensitive populations. The confidence limit explains experimental variability, which includes physical factors (food, water, housing, etc.), dosing measurements, and variability in the response of inbred animals. Is it possible to develop default margins of exposure that take into consideration the shape of the dose-response curve at the point of departure? More guidance is needed on applying margin of exposure analysis in risk assessments.

(7) Pharmacokinetics modeling – Pharmacokinetic modeling is a substantial improvement over the dose scaling method for estimating internal and delivered doses as well as for interspecies dose extrapolation. Constructing a model requires a very rich data set, which is available for only a few agents. Guidance for evaluating accuracy of the models used to estimate doses and guidance for validating pharmacokinetic models may speed up the process for using these models in risk assessments. Specific guidance or criteria could also aid investigators in conducting experiments used to construct models for risk assessments purposes.

(8) Benchmark dose approach to noncancer risk assessment – There are a number of issues related to the benchmark approach to noncancer risk assessment.

(a) Analysis of non-quantal data or transformation into a form easily incorporated into mathematical model – Quantal data expressed as incidence of lesion are easily incorporated, but data presented as increased severity of lesions require transformation. In addition, continuous data can be modeled, but should be evaluated in terms of an adverse and non-adverse categories. The magnitude of a change in enzymes or hematological parameters, the degree of severity in tissue lesions, or the magnitude of body weight decrements can affect data transformation. For example, should a 10% decrement in body weight, which is usually considered to be an adverse effect in animal models be considered an adverse effect for humans? Should the effects seen in animal studies be defined in terms of the degree of hazard to animals or the degree of hazard humans?

(b) Uncertainty factors – There are a number of issues concerning the application of uncertainty factors to the BMD to derive reference values. Should factors of 10 each for intraspecies and intraspecies (3 for inhalation exposures) as applied to NOAELs be applied to the BMD? BMD are based on LEDs, which are effect levels. Should a factor be applied to the BMD to extrapolate to a non-adverse effect level, as is done for LOAELs when deriving reference values?

(c) Validating and testing the BMD approach – A number of studies has compared the BMD to the corresponding NOAEL in a study, but is this a valid comparison? The benchmark approach is seen as an improvement over the NOAEL/LOAEL approach; nevertheless, a method for testing and validating this approach for deriving plausible noncancer risk values for humans should be developed. Deriving BMDs require more data. For most chemicals, sufficient data will not be available. Should the risk assessment be postponed until more data become available, or should the NOAEL/LOAEL approach be used as a default approach when data are not available to use the benchmark approach?

(9) Multiple endpoints and critical effects – Depending on the selection of doses or the type of study conducted, multiple endpoints may be evaluated in a single study. This is particularly true for neurotoxicity and developmental studies. In the case of neurotoxicity studies, behavioral, neurochemical, neurophysiological, and structural effects of a toxicant may be evaluated. Developmental studies may include fetal weight, fetal death, morphological variations and malformations, as well a functional toxicity, which could include neurotoxicity. If only the critical effect (observed at the lowest dose) is evaluated, important information may not be included in the assessment. Should the critical effect be used without considering the severity of the effect? For example, should a decrease in fetal weight be given the same weight as malformations? Should effects that would have a potentially greater impact on the quality of life or on society be given more weight (e.g., structural variations vs malformations or cognitive deficits vs change in nerve conduction)? Should effects such as cancer, which occur after many years of exposure, be given more weight than potentially serious developmental effects that are irreversible throughout life span of an individual? How multiple endpoints or different types of endpoints are handled is an important issue in the risk assessment process.

(10) Establishing guidance for deriving RfD/RfC based on short-term exposures – The manifestation of neurotoxic effects may not be dependent on a long-term duration of exposure, and the manifestation of other effects that may not be dependent on a long-term duration of exposure. For example, adaptation may occur of there may be no increase in the severity of an effect with continued exposure. Whether the NOAEL/LOAEL or benchmark approach is used, guidance should be developed for deriving reference values based on hazards not associated with long-term exposure.

7. Glossary

The following definitions were obtained from the sources presented in this report.

Adverse effect - a biochemical change, functional impairment, or pathological lesion that either singly or in combination that compromise the performance of the whole organism.

Components - points in a risk assessment at with judgments must be made regarding the analytic approach to be taken.

Critical effect - the adverse effect or the known precursor to the adverse effect that first appears in the dose scale as the dose is increased.

Inference Options/Default Options - choices made among several scientifically plausible options.

LOAEL (lowest-observed-adverse-effect level) -the lowest exposure level at which there are statistically or biologically significant increases in frequency or severity of adverse effects between the exposed population and its appropriate control group.

NOAEL (no-observed-adverse-effect level) - an exposure level at which there are no statistically or biologically significant increases in the frequency or severity of adverse effects between the exposed population and its appropriate control group. Effects may be produced at this level, but they are not considered to be adverse, nor precursors to specific adverse effects.

NOEL (no-observed-effect level) - an exposure level at which there are no statistically or biologically significant increases in the frequency or severity of any effect between the exposed population and its appropriate control group.

Risk Assessment - qualitative or quantitative characterization of the potential health effects of particular substances on individuals or populations.

fD or RfC (reference dose or reference concentration) - an estimate (with uncertainty spanning perhaps an order of magnitude) of the daily exposure to the human population (including sensitive subgroups) that is likely to be without appreciable risk of deleterious effects during a lifetime.

Risk Assessment Policy - choices made based on both science and broader areas of social and economic considerations.

Risk Management - process of evaluating alternative regulatory actions and selecting among them; entails political, social, economic, and technological considerations; and involves the use of value judgements related to acceptability of risk and reasonableness of cost.

Uncertainty factors - generally 10-fold factors (but may be less) representing specific areas of scientific uncertainty inherent in the extrapolation from the available data (sensitive subpopulations, species extrapolation, less-than-lifetime extrapolation, LOAEL to NOAEL extrapolation, and insufficient database).

8. REFERENCES

Allen, B.C.; Kavlock, R.J.; Kimmel, C.A.; Faustman, E.M. 1994. Dose-response assessment for developmental toxicity. II. Comparison of generic benchmark dose estimates with no observed adverse effect levels. Fund. Appl. Toxicol. 23:487-495.

Anderson, E.; Deisler, Jr., P.F.; McCallum, D.; et al. 1993. Key issues in carcinogen risk assessment guidelines, society for risk analysis. Risk. Anal. 13:379-382.

Annau, Z. 1990. Behavioral toxicology and risk assessment. Neurotoxicol. Teratol. 12:547-551.

Anonymous. 1993. Toxicity tests in animals: extrapolating to human risk. Environ. Health Perspect. (Suppl.) pp. 396-401.

Ashby, J.; Doerrer, N.G.; Flamm, F.G.; et al. 1990. A scheme for classifying carcinogens. Reg. Toxicol. Pharmacol. 12:270-295.

Barnes, D.G.; Daston, G.P.; Evans, J.S.; et al. 1995. Benchmark Dose Workshop: criteria for use of a benchmark dose to estimate a reference dose. Reg. Toxicol. Pharmacol. 21:296-306.

Beck, B.D.; Conolly, R.B.; Dourson, M.L.; et al. 1993. Symposium overview Improvements in quantitative noncancer risk assessment. Fund. Appl. Toxicol. 20:1-14.

Black, D.L.; Marks, T.A. 1992. Role of maternal toxicity in assessing developmental toxicity in animals: a discussion. Reg. Toxicol. Pharmacol. 16:189-201.

Butterworth, B.E.; Goldsworthy, T.L.; Popp, J.A.; McClellan, R.O. 1991. The rodent cancer test: as assay under siege. CIIT Activities. 11(9):1-8.

Calabrese, E.J. (Ed.) 1991. Principles of Animal Extrapolation. Lewis Publ., Chelsea, MI. PP. 499-574.

Calabrese, E.J.; Beck, B.D.; Chappell, W.R. 1992. Does the animal-to-human uncertainty factor incorporate interspecies differences in surface area? Reg. Toxicol. Pharmacol. 15:172-179.

Carr, C.J.; Kolbye, Jr., A.C. 1991. A critique of the use of the maximum tolerated dose in bioassays to assess cancer risks from chemicals. Reg. Toxicol. Pharmacol. 14:78-87.

Cohen, B.L. 1981. Long-term consequences of the linear-no-threshold dose-response relationship for chemical carcinogens. Risk Analysis 1:267-275.

Cohen, S.M.; Ellwein, L.B. 1992. Risk assessment based on high-dose animal exposure experiments. Chem. Res. Toxicol. 5:742-748.

Cohen, S.M.; Ellwein, L.B. 1993. Use of cell proliferation data in modeling urinary bladder carcinogenesis. Environ. Health Perspect. 101 (Suppl. 5):111-114.

Crump, K.S. 1984a. A new method for determining allowable daily intakes. Fund. Appl. Toxicol. 4:854-871.

Crump, K.S. 1984b. An improved procedure for low-dose carcinogenic risk assessment from animal data. J. Environ. Pathol. Toxicol. 5:339-348.

Daston, G.P.; Rogers, J.M.; Versteeg, D.J; et al. 1991. Interspecies comparisons of A/D ratios: A/D ratios are not constant across species. Fund Appl Toxicol 17:696-722.

Dews, P.B. 1986. On the assessment of risk. In: *Developmental Behavioral Pharmacology*, N. Krasnegor, J. Gray, and T. Thompson, Eds. , Lawrence Earlbaum Assoc. Hillsdale, NJ.

Dourson, M.L.; Stara, J.F. 1983. Regulatory history and experimental support of uncertainty (safety) factors. Reg. Toxicol. Pharmacol. 3:224-238.

Eastern Research Group, Inc. 1994. *Report on the Workshop on Cancer Risk Assessment Guidelines issues.* Risk Assessment Forum, September, 1994, U.S. Environmental Protection Agency, Washington, D.C.

Farland, W.; Dourson, M. 1993. Noncancer health endpoints: Approaches to quantitative risk assessment. In: *Comparative Environmental Risk Assessment*, C.R. Cothern, Ed., Lewis Publ., pp.87-106.

Francis, E.Z. 1992. Regulatory developmental neurotoxicity and human risk assessment. Neurotoxicol. 13:77-84.

Gad, S.C.; Weil, C.S., Eds. 1986. Carcinogenesis and Risk Assessment. In: *Statistics and Experimental Design for Toxicologist*, Telford Press, Caldwell, NJ. pp. 176-243.

Gaylor, D.W. 1983. The use of safety factors for controlling risk. J. Toxicol. Environ. Health. 11:329-336.

Gaylor, D.W.; Slikker, W., Jr. 1990. Risk assessment for neurotoxic effects. Neurotoxicology 11:211-218.

Glowa, J.R; Dews, P.B. 1987. Behavioral toxicology of volatile organic solvents. IV. Comparison of the behavioral effects of acetone, methyl ethyl ketone, ethyl acetate, carbon disulfide, and toluenes on the responding of mice. J. Am. Coll. Toxicol. 6:461-469.

Glowa, J.; DeWeese, J.; Natale, M.E.; Holland, J.J. 1983. Behavioral toxicology of volatile organic solvents. I. Methods: acute effects. J. Am. Coll. Toxicol. 2:175-185.

Goodman, G.; Wilson. R. 1991. Predicting the carcinogenicity of chemicals in humans from rodent bioassay data. Environ. Health Perspect. 94:195-218.

Gregory, A.R. 1988. Species comparisons in evaluating carcinogenicity in humans. Reg. Toxicol. Pharmacol. 8:160-190.

Griesemer, R.A. 1992. Dose selection for animal carcinogenicity studies: A practitioner's perspective. Chem. Res. Toxicol. 5:737-741.

Harvard Center for Risk Analysis. 1994. *A Historical Perspective on Risk Assessment in the Federal Government.* Harvard School of Public Health.

Haseman, J.K. 1985. Issues in carcinogenicity testing: Dose selection. Fund. Appl. Toxicol. 5:66-78.

Haseman, J.K.; Lockhart, A. 1994. The relationship between use of the maximum tolerated dose and study sensitivity for detecting rodent carcinogenicity. Fund. Appl. Toxicol. 22:382-391.

Huff, J. 1993. Absence of morphologic correlation between chemical toxicity and chemical carcinogenesis. Environ. Health Perspect. 101 (Suppl. 5):45-54.

ILSI (International Life Science Institute). 1996. In preparation.

Jarabek, A.M. 1995a. Interspecies extrapolation based on mechanistic determinants of chemical disposition. Human Eco. Risk Assess. 1:641-662.

Jarabek, A.M. 1995b. The application of dosimetry models to identify key processes and parameters for default dose-response assessment approaches. Toxicol. Lett. 79:171-184.

Jarabek, A.M.; Menache, Overton, Jr., J.H.; Dourson, M.L.; Miller, F.J. 1990. The U.S. Environmental Protection Agency's inhalation RfD methodology: risk assessment for air toxics. Toxicol. Ind. Health. 6:279-301.

Johannsen, F.R. 1990. Risk assessment of carcinogenic and noncarcinogenic chemicals. Crit. Rev. Toxicol. 20:341-366.

Johnson, E.M.; Gabel, B.E.G. 1983. An artificial embryo for detection of abnormal developmental biology. Fund. Appl. Toxicol. 3:243-249.

Khera, K.S. 1984. Maternal toxicity - a possible factor in fetal malformations in mice. Teratology. 29:411-416.

Kimmel, C.; Gaylor, D. 1988. Issues in qualitative and quantitative risk analysis for developmental toxicology. Risk Anal. 8:15-21. (cited in U.S. EPA, 1995)

Manson, J.M.; Wise, L.D. 1986. Teratogens. In: *Casarett and Doull's Toxicology: The Basic Science of Poisons.* 4th ed., M.O. Amdur, C.D. Klaassen, and J. Doull, Eds., McGraw-Hill, New York, NY. pp. 226-254.

Monticello, T.M.; Gross, E.A.; Morgan, K.T. 1993. Cell proliferation and nasal carcinogenesis. Environ. Health Perspect. 101 (Suppl. 5):121-124.

Moolgavkar, S.H. 1993. Cell proliferation and carcinogenesis models: general principles with illustrations from the rodent liver system. Environ. Health Perspect. 101 (Suppl. 5):91-94.

Munro, I.C. 1988. Qualitative factors in carcinogen classification. In: *Banbury 31: Carcinogen Risk Assessment: New Directions in the Qualitative and Quantitative Aspects.* Cold Spring Harbor Laboratory, NY. pp. 211-229.

NRC/NAS (National Research Council/National Academy of Sciences). 1983. *Risk Assessment in the Federal Government: Managing the Process.* National Academy Press, Washington, D.C.

NRC/NAS (National Research Council/National Academy of Sciences). 1994. *Science and Judgment in Risk Assessment.* Committee on Risk Assessment of Hazaradous Air Pollutants, Board on Environmental Studies and Toxicology, commission on Life Sciences, National Research Council, National Academy Press, Washington, D.C.

NTP (National Toxicology Program). 1992. Final report of the Advisory Review by the NTP Board of Scientific Counselors. Fed. Reg. 57:31721-21730.

OSTP (Office of Science and Technology Policy). 1985. Chemical carcinogens: reviews of the science and its associated principles. Fed. Reg. 50:10372-10442.

Pereira, M.A. 1985. Mouse liver tumor data: assessment of carcinogenic activity. Toxicol. Ind. Health. 1:311-333.

Purchase, I.F.H. 1994. Current knowledge of mechanisms of carcinogenicity: Genotoxins versus non-genotoxins. Human Exp. Toxicol. 13:17-28.

Rees, D.C.: Glowa, J.R. 1994. Extrapolation to humans for neurotoxicants. In: *The Vulnerable Brain and Environmental Risks*, Vol. 3: Toxins and Water, R.L. Isaacson and K.F. Jensen, Eds. Plenum Press, New York.

Reiter, L.W. 1987. Neurotoxicology in regulation and risk assessment. Devel. Pharmacol. Ther. 10:354-368.

Reynolds, S.H.; Stowers, S.J.; Patterson, R.M.; Maronpot, R.R.; Anderson, M.W. 1988. Oncogene activation in spontaneous and chemically induced rodent tumors: implications for risk analysis. Environ. Health Perspect. 78:175-177.

Shoaf, C.R. 1991. Current assessment practices for noncancer end points. Environ. Health Perspect. 95:111-119.

Short, B.G. 1993. Cell proliferation and renal carcinogenesis. Environ. Health Perspect. 101(Suppl. 5):115-120.

Sontag, J.M.; Page, N.P.; Safiotti, U. 1976. *Guidelines for Carcinogen Bioassay in Small Rodents.* DHHS Publ. (NIH) 76-801. National Cancer Institute, Bethesda, MD.

Spencer, P.S.; Schaumberg, H.H. 1985. Organic solvent neurotoxicity: Facts and research. Scand. J. Work Environ. Health 11:53-60. (cited in Rees and Glowa, 1994)

Stanton, M.E.; Spear, L.P. 1990. Workshop on the qualitative and quantitative comparability of human and animal developmental neurotoxicity. Work Group Report: Comparability of measures of developmental neurotoxicity in humans and laboratory animals. Neurotoxicol. Teratol. 12:261-267. (cited in Winneke, 1992)

Tilson, H.A. 1990. Neurotoxicology in the 1990s. Neurotoxicol. Teratol. 12:293-300.

Tilson, H.A.; Cabe, P.A. 1978. Strategy for the assessment of neurobehavioral consequences of environmental factors. Environ. Health Perspect. 26:287-299. (cited in Reiter, 1992)

Travis, C.C.; White, R.K. 1988) Interspecies scaling of toxicity data. Risk Anal. 8:119-125.

Tyl, R.W.; Sette, W.F. 1990. Workshop on the qualitative and quantitative comparability of human and animal developmental neurotoxicity, Work Group III report: weight of evidence and quantitative evaluation of developmental neurotoxicity data. Neurotoxicol. Teratol. 12:275-280.

U.S. EPA. 1980. Guidelines and methodology used in the preparation of health effect assessment chapters of the consent decree water criteria documents. Fed. Reg. 45:79347-79379.

U.S. EPA. 1985. Toxic substances control act test guidelines. 40CFR 798. Subpart G - neurotoxicity. Section 798.6500. Fed. Reg. 50:39465-39466.

U.S. EPA. 1986. Guidelines for carcinogen risk assessment. Fed. Regist. 51:33992-34003.

U.S. EPA. 1988a. Proposed guidelines for assessing male reproductive risk and request for comments. Fed. Reg. 53:24850-24869.

U.S. EPA. 1988b) Proposed guidelines for assessing female reproductive risk; notice. Fed. Reg. 53:24834-24847.

U.S. EPA, 1989a. *Interim Methods for Development of Inhalation Reference Doses.* Environmental Criteria and Assessment Office, Office of Health and Environmental Assessment, Research Triangle Park, NC. EPA/600/8-88/066F

U.S. EPA. 1989b. *Workshop Report on EPA Guidelines for Carcinogen Risk Assessment: Use of Human Evidence.* Assembled by Eastern Res. Group, Inc. for the Risk Assessment Forum, Technical panel on Carcinogen Guidelines. EPA/625/3-90/017.

U.S. EPA. 1991a. *Alpha 2_U-Globulin: Association with Chemically Induced Renal Toxicity and Neoplasia in the Male Rat.* Risk Assessment Forum, Washington, D.C. EPA/625/3-91/019F.

U.S. EPA. 1991b. *General Quantitative Risk Assessment Guidelines for Noncancer Health Effects.* External Review Draft. Environmental Criteria and Assessment Office, Office of Health and Environmental Assessment, Cincinnati, OH. ECAO-CIN-538.

U.S. EPA. 1991c. Guidelines for developmental toxicity rsk asessment. Fed. Reg. 56:63799-63826.

U.S. EPA. 1991d. *Pesticide assessment guidelines, Subdivision F, Hazard Evaluation, Neurotoxicity.* Health Effects Division, Office of Pesticide Programs. EPA 540/09-91-123.

U.S. EPA 1992a. *Working Paper for Considering Draft Revisions to the U.S. EPA Guidelines for Cancer Risk Assessment.* Office of Health and Environmental Assessment, Office of Research and Development, Washington, D.C. EPA/600/AP-92/003.

U.S. EPA. 1992b. Draft report: A cross-species scaling factor for carcinogen risk assessment based on equivalence of $mg/kg^{3/4}$/day. Fed. Reg. 57:24152-24173.

U.S. EPA. 1993. *Reference Dose (RfD): Description and Use in Health Risk Assessments.* IRIS (Integrated Risk Information System) Database, Environmental Criteria and Assessment Office, Office of Health and Environmental Assessment, Cincinnati, OH.

U.S. EPA. 1994a. *Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry.* Environmental Criteria and Assessment Office, Office of Health and Environmental Assessment, Research Triangle Park, NC. EPA/600/8-90/066F

U.S. EPA. 1994b. Final report: principles of neurotoxicity risk assessment. Fed. Reg. 59:42360-42404.

U.S. EPA. 1995. *The Use of the Benchmark Dose Approach in Health Risk Assessment.* Risk Assessment Forum, Office of Research and Development, Washington, D.C. EPA/630/R-94/007

U.S. EPA. 1996a. *PCBs: Cancer Dose-Response Assessment and Application to Environmental Mixtures.* National Center for Environmental Assessment, Office of Research and Development, Washington, D.C. EPA/600/P-96/001A.

U.S. EPA. 1996b. *Proposed Guidelines for Carcinogen Risk Assessment.* Office of Research and Development, Washington, D.C. EPA/600/P92/003C.

Ward, J.M.; Uno, H.; Kurata, Y.; Weghorst, C.M.; Jang, J.-J. 1993. Cell proliferation not associated with carcinogenesis in rodents and humans. Environ. Health Perspect. 101 (Suppl. 5):125-136.

Weil, C.S. 1972. Statistics versus safety factors and scientific judgement in the evaluation of safety for man. Toxicol. Appl. Pharmacol. 21:454-463. (cited by Dourson and Stara, 1983)

Wilkinson, C.F. 1987. Being more realistic about chemical carcinogenesis. Environ. Sci. Technol. 21:843-847.

Winneke, G. 1992. Cross species extrapolation in neurotoxicology: Neurophysiological and neurobehavioral aspects. Neurotoxicology 13:15-26.

Woo, D.C.; Hoar, R.M. 1972. "Apparent hydronephrosis" as a normal aspect of renal development in late gestation of rats: the effect of methyl salicylate. Teratology 6:191-196.